PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Pater		A 1	(1	1) International Publication Number: WO 98/53061	
C12N 15/12, C07K 14/47, 16/18, G01N 33/53		A1	(4	3) International Publication Date: 26 November 1998 (26.11.98)	
(21) International Application Number: PCT/AU98/00380			80	(74) Agents: HUGHES, E., John, L. et al.; Davies Collison Cave, Little Collins Street, Melbourne, VIC 3000 (AU).	
(22) International Filin	g Date: 22 May 1998 (22.05.9	(8)		
(30) Priority Data:	22 May 1007 (22 05 07)			(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, CM, CM, CM, CM, LM, LB, LS, LB, KE, KE, KB, KB, KB, KB, KB, KB, KB, KB, KB, KB	
PO 6972 PO 6973	23 May 1997 (23.05.97) 23 May 1997 (23.05.97)		U U	GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW,	
PO 6973	23 May 1997 (23.05.97)		เบ	MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL,	
PP 1458	22 January 1998 (22.01.98)		เบ	TJ. TM. TR. TT. UA. UG. US. UZ. VN. YU. ZW. ARIPO	
PP 1459	22 January 1998 (22.01.98)		Ū	patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian	
PP 1460	22 January 1998 (22.01.98)		ΔU	patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF,	
CIL OF THE Q	designated States except US): THE UEENSLAND INSTITUTE OF M .U/AU]; 300 Herston Road, Brisba	IEDIC/	۱L	CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
4029 (AU).				Published	
	•			With international search report.	
(72) Inventors; and					
[AU/AU]; 13 P SILINS, Ginters QLD 4061 (AU Research Counc	nts (for US only): HAYWARD, rince Street, Paddington, QLD 40 s [AU/AU]; 35 Eppalong Street, J). GRIMMOND, Sean [AU/GB]; il, Harwell, Didcot, Oxfordshire ODE, Michael [AU/AU]; 19 Thom	064 (AU The Ga Media X11 OF	U). ap, cal SD		
Camp Hill, QLI	D 4152 (AU). HANCOCK, John [d, Pullenvale, QLD 4069 (AU).				

(54) Title: THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN

(57) Abstract

The present invention relates generally to three novel human genes with gene regulatory function. These genes encode a zinc finger protein, a guanine nucleotide exchange protein and a heat shock protein or heat shock binding protein. The invention includes derivatives and mammalian animal, insect, nematodes, avian and microbial homologues of these genes. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL AM AT AU AZ BB BB BB BF BG BJ BR CCF CG CH CI CM CU CZ DE DK EE	Albania Armenia Austria Australia Azerbaijan Bosnia and Herzegovina Barbados Belgium Burkina Faso Bulgaria Benin Brazil Belarus Canada Central African Republic Congo Switzerland Côte d'Ivoire Cameroon China Cuba Czech Republic Germany Denmark Estonia	ES FI FR GA GB GE GH GN IE IL IS IT JP KE KG KP KR KZ LC LI LK LR	Spain Finland France Gabon United Kingdom Georgia Ghana Guinea Greece Hungary Ireland Israel Iceland Italy Japan Kenya Kyrgyzstan Democratic People's Republic of Korea Republic of Korea Republic of Korea Kazakstan Saint Lucia Liechtenstein Sri Lanka Liberia	LS LT LU LV MC MD MG MK ML MN MR MW MX NE NL NO NZ PL PT RO SE SG	Lesotho Lithuania Luxembourg Latvia Monaco Republic of Moldova Madagascar The former Yugoslav Republic of Macedonia Mali Mongolia Mauritania Malawi Mexico Niger Netherlands Norway New Zealand Poland Portugal Romania Russian Federation Sudan Sweden Singapore	SI SK SN SZ TD TG TJ TM TR TT UA UG US VN YU ZW	Slovenia Slovakia Senegal Swaziland Chad Togo Tajikistan Turkmenistan Turkey Trinidad and Tobago Ukraine Uganda United States of America Uzbekistan Viet Nam Yugoslavia Zimbabwe
--------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

WO 98/53061

-1-

THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN

PCT/AU98/00380

FIELD OF THE INVENTION

5 The present invention relates generally to a novel human gene and its derivatives and to mammalian, animal, insect, nematodes, avian and microbial homologues thereof. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

10

BACKGROUND OF THE INVENTION

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

15

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. There is growing need to develop recombinant and genetic molecules for use in diagnosis and in conventional pharmaceutical preparations as well as in gene and protein replacement therapies.

20

In work leading up to the present invention, the inventors sought to identify and clone human genes which might be useful as potential diagnostic and/or therapeutic agents. Molecules of particular interest targeted by the inventors were gene regulators including regulatory proteins, signal transducters and heat shock proteins.

25

Gene expression generally requires interaction between a regulatory protein and an appropriate recognition sequence of a target gene. Regulatory proteins comprise in many cases a domain or motif which facilitates binding to DNA. One particular motif comprises small sequence units repeated in tandem with each unit folded about a zinc atom to form separate structural domains.

30 This motif is now referred to as a zinc finger domain. Such a domain is generally defined by the number of cysteine (C) and histidine (H) residues.

In addition, knowledge of cellular interaction in the control of cell proliferation is essential in the rational design of specific therapeutic strategies aimed at controlling proliferative disorders. Such proliferative disorders including a range of cancers, inflammatory conditions and atherosclerosis. An important aspect of cellular interaction is in signal transduction via receptors to intracellular transducers. One key signal transducer is Ras which couples the receptors for diverse extracellular signals to different effectors. Ras directly activates the downstream kinase Raf which in turn induces the mitogen activated protein kinase (MAPK) cascade.

Another regulatory mechanism involves heat shock proteins. The *Escherichia coli* heat shock protein, DnaJ, is the founding member of a family of proteins which are associated with protein folding, protein complex assembly and transit through subcellular components.

Prokaryotic and eukaryotic DnaJ homologues have a modular organisation consisting of a J domain, a glycine-rich spacer, CXXCXGXG [SEQ ID NO:1] repeats and a C-terminal region with no obvious sequence features, as well as additional sequences for protein targeting. The J domain is anticipated to mediate interaction with heat shock 70 proteins (Hsp70) and consists of some 70 amino acids, frequently located at the N-terminus of the protein.

In accordance with the present invention, a genes have been identified from the human genome which encodes proteins having a regulatory role. One gene, in accordance with the present invention encodes a protein with an N-terminal region resembling a zinc-finger domain of a novel type. Another gene encodes a protein involved in guanine nucleotide exchange factor (GEF) signalling pathways. Yet another gene encodes a protein which is a heat shock protein or heat shock-like protein which may have a role in tumour suppression.

25

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Sequence identity numbers (SEQ ID NOs.) for nucleotide and amino acid sequences referred to in the subject specification are defined after the bibliography. A summary of SEQ ID NOs. is also given in Table 1.

- 5 One aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
- 10 Another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an (HC₁)₂ type.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:2 defines the gene, mcg4. This gene encodes a product, MCG4, having an amino acid sequence set forth in SEQ ID NO:3.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg4 gene portion, which mcg4 gene portion is capable of encoding an MCG4 polypeptide or a 30 functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg4 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.

A further aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an 20 amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

25

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the

nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:4 or 6 defines the gene, mcg7. This gene encodes a product, MCG7, having an amino acid sequence set forth in SEQ ID NO:5 or 7.

5

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg7 gene portion, which mcg7 gene portion is capable of encoding an MCG7 polypeptide or a functional or immunologically interactive derivative thereof.

10

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg7 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock binding protein or a derivative thereof.

Another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 5 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
 - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 41°C to the nucleotide sequence set forth in (i), (ii) or (iii).

10

The nucleotide sequence set forth in SEQ ID NO:8 defines the gene, mcg18. This gene encodes a product, MCG18, having an amino acid sequence set forth in SEQ ID NO:7.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg18 gene portion, which mcg18 gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg18 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

25

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

30

Another aspect of the present invention contemplates a method for detecting MCG18 or a

WO 98/53061 PCT/AU98/00380

-7-

derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

5

A summary of SEQ ID Nos. referred to in the subject specification is shown in Table 1.

TABLE 1
SUMMARY OF SEQ ID Nos.

5	SEQ ID NO.	DESCRIPTION		
1		amino acid repeat sequence in DnaJ homologues		
	2	Nucleotide sequence of mcg4		
	3	amino acid sequence of MCG4		
	4	nucleotide sequence of mcg7		
10	5	amino acid sequence of MCG7		
	6	nucleotide sequence of mcg7 within exon of		
		nucleotides 183-288		
	7	amino acid sequence of MCG7 within exon of		
		nucleotide 183-288		
	8	nucleotide sequence of mcg18		
	9	amino acid sequence of MCG18		
15	10-18	amino acid sequence identified using BESTFIT		
	19	sequence of pGEX and mcg7 junction		
	20	sequence of pGEX and mcg7 junction		
	21	nucleotide sequence of myc-tag/mcg7 junction		
	22	amino acid sequence corresponding to SEQ ID NO:21		
20	23	nucleotide sequence of pGEX and mcg7 junction		
	24	amino acid sequence corresponding to SEQ ID NO:23		
	25-36	mcg7-specific oligonucleotide		
	37-45	mcg18-specific oligonucleotide		

²⁵ Single and three letter abbreviations for amino acid residues are shown in Table 2.

- 9 -

TABLE 2

Amino Acid	Three-letter	One-letter
·	Abbreviation	Symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	С
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	Н
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	w
Tyrosine	Tyr	Υ.
Valine	Val	v
Any residue	Xaa	X

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation of the nucleotide sequence [SEQ ID NO:2] and corresponding amino acid sequence [SEQ ID NO:3] of mcg4.

5

Figure 2 is a representation of the alignment of the human MCG4 amino acid sequence with a translation of a partial murine expressed sequence tag (EST).

Figure 3 is a representation of the alignment of the human MCG4 amino acid sequence with a 10 translation of a partial nematode EST.

Figure 4 is a diagrammatic representation showing a predicted structure of MCG4 where H and C represent histidine and cysteine residues, respectively and X refers to any amino acid residue. Zn represent zinc atoms.

15

Figure 5 is a representation of sensitive sequence homology search of related cysteine-containing motifs in another *Caenorhabditis elegans* protein.

Figure 6 is a representation showing that a related cysteine containing motif is present in the 20 GATA-binding transcription factor from Saccharomyces pombe.

Figure 7 is a Northern blot showing expression of mcg4 in various cultured human cancer cell lines. Lanes 1-5, respectively, represent the hybridization signal from 15μg total RNA derived from various human cancer cell lines. Lanes 1-5, respectively, contain RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

Figure 8 is a representation of a partial alignment of *mcg4* with human ESTs AA074703 and AA134788.

30

Figure 9 is a representation of the partial nucleotide sequence alignment between a human

WO 98/53061 PCT/AU98/00380

- 11 -

(W32939) and mouse (AA242159) mcg4-like EST in the putative 5' UTR of the mcg4 cDNA. The putative initiation codon is underlined and the region upstream represents 5' UTR.

Figure 10 is a representation showing MacVector alignment of MCG4 with forward translations of ESTs AA134788 and AA074703. The nucleotide sequences are shown in Figure 8.

Figure 11 is a diagrammatic representation of the domains of MCG4

zinc finger consensus: CX₂HX₄CX₂CX₄HX₂CX₁₇CX₂CX₁₈HX₂CX₁₈CX₂C

acidic domain consensus: 9/34 amino acids negatively charged, 0/34 positively charged

basic domain consensus: 13/55 amino acids positively charged, 0/55 negatively charged

leucine zipper domain consensus: LX₆LX₆RX₆LX₆L

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha he!ix domain: (aa261) LX₆LXLX₆LXLX₆L (aa 286).

15 Figure 12 is a representation showing similarity of MCG7 with GEFs of various organisms.

Figure 13(a) is a representation of the nucleotide sequence [SEQ ID NO:4] and corresponding amino acid sequence [SEQ ID NO:5] of mcg7. Nucleotides 183-288 are an alternative spliced exon (shown in lower case).

20

Figure 13(b) is a representation of the partial nucleotide sequence [SEQ ID NO:6] and corresponding amino acid sequence [SEQ ID NO:7] of mcg7 but without the exon shown in Fig. 13(a). Amino acids have been numbered from the first methionine codon (underlined). The cDNA molecules of Fig. 13(a) and Fig. 13(b) differ by the inclusion and exclusion of the exon 25 of nucleotides 183-288.

Figure 14 is a representation showing a comparison between MCG7 and a homologue from Caenorhabditis elegans using the BESTFIT algorithm. In the figure, the following sequences are underlined:

30

1a nematode DVDEEDEVEDIEF [SEQ ID NO:10]

1b human DVDGDGHISQEEF [SEQ ID NO:11]

nematode DHDRDGFISQEEF [SEQ ID NO:12]

1c human DQNQDGCISREEM [SEQ ID NO:13]

nematode DVDMDGQISKDEL [SEQ ID NO:14]

GUANINE NT BINDING REGION = BLOCKS DATABASE NO. BL00720B

2 human

5

HFVHVAEKLLQLQNFNTLMAVVGGLSHSSISRLKETH[SEQID NO:15]

nematode

KFVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTY

10 [SEQ ID NO:16]

DaG-PE BINDING DOMAIN = PROSITE DATABASE NO. PD0C00379

3 human HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC [SEQ ID NO:17]

15 nematode HNFHETTFLTPTTCNHCNKLLWGILRQGFKCKDCGLAVHSCCKSNAVAEC [SEQ ID NO:18]

Figure 15 is a representation of an alignment of human and a partial (5' UTR and partial coding sequence) murine mcg7 cDNA (GenBank Acc. No. W71787 and AA237373). The putative initiation codon is underlined. The murine sequence represents a composite of 2 partial cDNA sequences from the EST database (accession numbers W71787 and AA237373). Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

- 25 Figure 16 is a representation of further 5' nucleotide and corresponding amino acid sequence for human mcg7. Nucleotide positions 1-321 were derived from GenBank Acc. No. AC000134 and nucleotides 322 onwards from Fig. 13(a). Two in-frame initiation codons are underlined. Asterisks denote in-frame stop codons.
- 30 Figure 17 is a graphical representation of a GDP release assay. □ Experiment #1 (mean of duplicates). ♦ Experiment #2 (mean of duplicates). The exchange reaction contained 36pmols

of GST-MCG (N-terminally truncated; encoded by Construct B in Fig. 18) and 1.6-12.8 pmols of recombinant GST-N-Ras.GDP. Reaction time 6 mins.

Estimated reaction constants:

 $K_m = 2.1 \mu M$, $V_{max} = 37 p Mol/6 min/36 p Mol [Expt#1]$

5 $K_m = 1.5 \mu M$, $V_{max} = 30.3 p Mol/6 min/36 p Mol [Expt#2]$

Figure 18 depicts various recombinant plasmids containing partial or full-length mcg7.

Figure 19 is a representation of the nucleotide sequence [SEQ ID NO:8] and corresponding amino acid sequence [SEQ ID NO:9] of mcg18.

Figure 20 is a representation showing that MCG18 has partial homology to E. coli DnaJ.

Figure 21 is a representation showing that MCG18 has homology to two *Caenorhabitis elegans* 15 proteins.

Figure 22 is a representation showing that MCG18 has homology to a Saccharomyces pombe protein.

20 Figure 23 is a representation showing homology of MCG18 to a Drosophila virilis protein.

Figure 24 is a representation showing homology of MCG18 to human DnaJ proteins HDJ-2/HSDJ, HDJ-1/HSP40 and HSJ1.

25 Figure 25 is a representation of the nucleotide and corresponding amino acid sequence of murine mcg18.

Figure 26 is a representation of homology between human and murine MCG18.

30 Figure 27 depicts nucleotide sequences corresponding to the 5' untranslated region of human mcg18.

WO 98/53061 PCT/AU98/00380

- 14 -

Figure 28 depicts a Northern blot showing expression of mcg18 transcripts in total RNA isolated from various human cancer cell lines grown in culture. Lanes 1-5 respectively contain $15\mu g$ RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

5

25

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having 5 homology to a regulator of gene expression or a derivative of said gene regulator.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an (HC₃)₂ type.

Still more particularly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- 15 (i) a nucleotide sequence set forth in SEO ID NO:2;
 - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
 - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The present invention also provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- 30 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;

- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

5

Another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock-binding protein or a derivative thereof.

10

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 15 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
 - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

20

Preferably, the percentage similarity is at least about 50%. More preferably, the percentage similarity is at least about 60%.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for hybridisation, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least

about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels.

The present invention extends to nucleic acid molecules with percentage similarities of approximately 65%, 70%, 75%, 80%, 85%, 90% or 95% or above or a percentage in between.

15 The nucleic acid molecule of the present invention defined by SEQ ID NO:2 is hereinafter referred to as constituting the "mcg4" gene. The protein encoded by mcg4 is referred to herein as "MCG4" and has an amino acid sequence set forth in SEQ ID NO:3. The mcg4 gene is proposed to encode, in accordance with the present invention, a regulator of gene expression and comprises a novel zinc finger domain, (HC₃)₂. A regulator of gene expression includes a transcription factor. Regulation may be at the level of nucleic acid:protein or protein:protein interaction.

The nucleic acid molecule of the present invention defined by SEQ ID NO:4 or 6 is hereinafter referred to as constituting the "mcg7" gene. The protein encoded by mcg7 is referred to herein as "MCG7" and has an amino acid sequence set forth in SEQ ID NO:5 or 7 and is involved in signal transduction. The difference in the nucleotide and amino acid sequence is due to the presence or absence of an exon at nucleotides 183-288.

The nucleic acid molecule of the present invention defined by SEQ ID NO:8 is hereinafter 30 referred to as constituting the "mcg18" gene. The protein encoded by mcg18 is referred to herein as "MCG18" and comprises the amino acid set forth in SEQ ID NO:9.

15

The present invention extends to the naturally occurring genomic mcg4, mcg7 and mcg18 nucleotide sequences or corresponding cDNA sequences or to derivatives thereof. Derivatives contemplated in the present invention include fragments, parts, portions, mutants, homologues and analogues of MCG4, MCG7 or MCG8 or the corresponding genetic sequences. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG4, MCG7 or MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to mcg4, mcg7 or mcg18. "Additions" to the amino acid or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG4" or "mcg4", "MCG7" or "mcg7" or "MCG8" or mcg18" includes reference to all derivatives thereof including functional derivatives and immunologically interactive derivatives of MCG4, MCG7 or MCG18.

The mcg4, mcg7 and mcg18 of the present invention are particularly exemplified herein from humans and in particular from human chromosome 11q13.

The present invention extends, however, to a range of homologues from, for example, primates, livestock animals (eg. sheep, cows, horses, donkeys, pigs), companion animals (eg. dogs, cats) laboratory test animals (eg. rabbits, mice, rats, guinea pigs), reptiles, birds (eg. chickens, ducks, geese, parrots), insects, nematodes, eukaryotic microorganisms and captive wild animals (eg. deer, foxes, kangaroos). Reference herein to mcg4 and mcg18 or their respective proteins MCG4, MCG7 and MCG18 includes reference to these molecules of human origin as well as novel forms of non-human origin.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or

both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include E. coli, Bacillus sp and Pseudomonas sp. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

- 5 Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg4 gene portion, which mcg4 gene portion is capable of encoding an MCG4 polypeptide or a functional or immunologically interactive derivative thereof.
- 10 Preferably, the mcg4 gene portion of the genetic construct is operably linked to a promoter in the vector such that said promoter is capable of directing expression of said mcg4 gene portion in an appropriate cell.

In addition, the *mcg4* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-Stransferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

20

temporal regulation of particular genes.

It is proposed in accordance with the present invention that MCG4 is a transcription factor involved in gene regulation. Mutations in mcg4 may result in aberrations in gene regulation leading to the development of or a propensity to develop various types of cancer. In this regard, although not wishing to limit the present invention to any one hypothesis or mode of action, it.

25 is proposed that mcg4 or its expression product may be involved in the tissue-specific or

A deletion or aberration in the mcg4 gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a 30 heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may

be determined by assaying for aberrations in the parents and/or proband of a subject under investigation.

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg4 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

10

Another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg7 gene portion, which mcg7 gene portion is capable of encoding an mcg7 polypeptide or a functional or immunologically interactive derivative thereof.

15

Preferably, the mcg7 gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said mcg7 gene portion in an appropriate cell.

20 In addition, the mcg7 gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-Stransferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG7 is a GEF involved in signal transduction. Mutations in mcg7 or MCG7 may result in defective control of cell proliferation leading to the development of or a propensity to develop various types of cancer.

30

A deletion or aberration in the mcg7 gene may also be important in the detection of cancer or

a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents of a subject under investigation.

5

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg7 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Yet another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human 15 mcg18 gene portion, which mcg18 gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the mcg18 gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said mcg18 gene portion 20 in an appropriate cell.

In addition, the *mcg18* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

25

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG18 is a transcription factor 30 involved in protein folding, protein complex assembly and transit through subcellular compartments. MCG18 may also have a role in tumour suppression. Thus mutations in mcg18

may result in the development of or a propensity to develop various types of cancer.

A deletion or aberration in the *mcg18* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a 5 heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents and/or proband of the subject under investigation.

10 According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg18 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or 15 a propensity to develop said condition.

The nucleotide substitutions, additions or deletions may be detected by any convenient means including nucleotide sequencing, restriction fragment length polymorphism (RFLP), polymerase chain reaction (PCR), oligonucleotide hybridization and single stranded conformation polymorphism analysis (SSCP) amongst many others. An aberration includes modification to existing nucleotides such as to modify glycosylation signal amongst other effects.

In an alternative method, aberrations in the mcg4, mcg7 and mcg18 genes are detected by screening for mutations in MCG4, MCG7 and MCG18, respectively.

25

A mutation in MCG4, MCG7 or MCG18 may be a single or multiple amino acid substitution, addition and/or deletion. The mutation in mcg4, mcg7 or mcg18 may also result in either no translation product being produced or a product in truncated form. A mutant may also be an altered glycosylation pattern or the introduction of side chain modifications to amino acid residues.

According to this aspect of the present invention, there is provided a method of detecting a condition caused or facilitated by an aberration in mcg4, mcg7 or mcg18 said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4, MCG7 or MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

A particularly convenient means of detecting a mutation in MCG4, MCG7 or MCG18 is by use of antibodies.

- 10 Accordingly another aspect of the present invention is directed to antibodies to MCG4, MCG7 or MCG18 and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to MCG4, MCG7 or MCG18 or may be specifically raised to MCG4, MCG7 or MCG18 or derivatives thereof. In the case of the latter, MCG4, MCG7 or MCG18 or their derivatives may first need to be associated with a carrier molecule.
- 15 The antibodies to MCG4, MCG7 or MCG18 of the present invention are particularly useful as diagnostic agents.

For example, antibodies to MCG4, MCG7 or MCG18 and their derivatives can be used to screen for wild-type MCG4, MCG7 or MCG18 or for mutated MCG4, MCG7 or MCG18 molecules.

20 The latter may occur, for example, during or prior to certain cancer development. A differential binding assay is also particularly useful. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of normal MCG4, MCG7 or MCG18 levels or the presence of wild-type MCG4, MCG7 or MCG18 may be important for diagnosis of certain cancers or a predisposition for development of cancers or for monitoring 25 certain therapeutic protocols.

As stated above antibodies to MCG4, MCG7 or MCG18 of the present invention may be monoclonal or polyclonal or may be fragments of antibodies such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

For example, specific antibodies can be used to screen for wild-type MCG4, MCG7 or MCG18 molecule or specific mutant molecules such as molecules having a certain deletion. This would be important, for example, as a means for screening for levels of MCG4, MCG7 or MCG18 in a cell extract or other biological fluid or purifying MCG4, MCG7 or MCG18 made by recombinant means from culture supernatant fluid or purified from a cell extract. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of wild-type MCG4, MCG7 or MCG18 or to a specific mutant phenotype or to a deleted or otherwise altered region.

15

Both polyclonal and monoclonal antibodies are obtainable by immunization of a suitable animal or bird with MCG4, MCG7 or MCG18 or its derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal or bird with an effective amount of MCG4, MCG7 or MCG18 or antigenic parts thereof or derivatives thereof, collecting serum from the animal or bird, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

25

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates a method for detecting MCG4, MCG7 or MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4, MCG7 or MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4, MCG7 or MCG18 complex to form, and then detecting said complex.

Preferably, the biological sample is a cell extract from a human or other animal or a bird.

The presence of MCG4, MCG7 or MCG18 may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4, 424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

15

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into 20 contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigenlabelled antibody. Any unreacted material is washed away, and the presence of the antigen is 25 determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both samrile and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the 30 art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain MCG4, MCG7 or MCG18 including cell extract

WO 98/53061 PCT/AU98/00380

or tissue biopsy. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

In the typical forward sandwich assay, a first antibody having specificity for the MCG4, MCG7

5 or MCG18 or an antigenic part thereof or a derivative thereof or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from room temperature to 37°C) to allow binding of any subunit present in the 15 antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

- 20 An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.
- By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-30 bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide

containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled 5 artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, betagalactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a 10 fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated. 15 usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

30

As stated above, the present invention extends to genetic constructs capable of encoding MCG4.

MCG7 or MCG18 or functional derivatives thereof. Such genetic constructs are also contemplated to be useful in modulating expression of specific genes in which mcg4, mcg7 or mcg18 is involved in tissue-specific or temporal regulation.

- 5 Accordingly, another aspect of the present invention is directed to a genetic construct comprising a nucleotide sequence encoding a peptide, polypeptide or protein and mcg4, mcg7 or mcg18 or a functional derivative or homologue thereof capable of modulating the expression of said nucleotide sequence.
- 10 As stated above, MCG18 is proposed to have a role in tumour suppression. Accordingly, it is further proposed in accordance with the present invention to use recombinant MCG18 in pharmaceutical preparations for treating arresting or otherwise ameliorating the effects of certain cancers.
- 15 Accordingly, another aspect of the present invention contemplates a method for treating, arresting or otherwise ameliorating the effects of a cancer in an animal or bird, said method comprising administering to said animal or bird an effective amount of MCG18 or a functional derivative thereof for a time and under conditions sufficient to treat, arrest or otherwise ameliorate the effects of said cancer.

20

The present invention, therefore, contemplates a pharmaceutical composition comprising MCG18 or a derivative thereof or a modulator of mcg18 expression or MCG18 activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to hereinafter as the "active ingredients". The active ingredients may also include anti-cancer agents or agents which facilitate actions of MCG18.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) and sterile powders for the extemporaneous preparation of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier may be a solvent medium containing, for example, water, ethanol, polyol (for example,

glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as licithin and by the use of superfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimersal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with 20 the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 25 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 µg and 2000 mg of active compound.

30 The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter. A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium

phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

The present invention also extends to forms suitable for topical application such as creams, lotions and gels.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known 20 in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the

treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

The principal active ingredient is compounded for convenient and effective administration in 5 effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 µg to about 2000 mg. Expressed in proportions, the active compound is generally present in from about 0.5 µg to about 2000 mg/ml of carrier. In the case of compositions containing supplementary active ingredients, the dosages are determined by reference to the usual dose and manner of administration of the said ingredients.

Effective amounts contemplated by the present invention include those amounts effective to ameliorate a condition. For example, it is envisaged that effective amounts would range from about $0.001~\mu g/kg$ body weight to about 100~mg/kg body weight. Alternatively, effective amounts of about $0.01~\mu g/kg$ body weight to about 10~mg/kg body weight or even $0.1~\mu g/kg$ body weight to about 1~mg/kg body weight. Administration may be per minute, hour, day, week, month or year or may only be a once off administration.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating mcg18 expression or MCG18 activity. The vector may, for example, be a viral vector.

As stated above, the present invention further contemplates a range of derivatives of MCG18. Derivatives include fragments, parts, portions, mutants, homologues and analogues of the MCG18 polypeptide and corresponding genetic sequence. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to the genetic sequence encoding MCG18. "Additions" to amino acid sequences or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG18" includes reference to all derivatives thereof including functional derivatives or MCG18 immunologically interactive derivatives.

Analogues of MCG18 contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

5

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with NaBH₄; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with NaBH₄.

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

20 Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide c: sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form 30 a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids, contemplated herein is shown in Table 3.

TABLE 3

	Non-conventional arnino acid	Code	Non-conventional amino acid	Code
5	α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
	α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
	aminocyclopropane-	Cpro	L-N-methylasparagine	Nmasn
	carboxylate		L-N-methylaspartic acid	Nmasp
10	aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
	aminonorbornyl-	Norb	L-N-methylglutamine	Nmgln
	carboxylate		L-N-methylglutamic acid	Nmglu
	cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
	cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
15	D-alanine	Dal	L-N-methylleucine	Nmleu
	D-arginine	Darg	L-N-methyllysine	Nmlys
	D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
	D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
	D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
20	D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
	D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
	D-isoleucine	Dile	L-N-methylproline	Nmpro
	D-leucine	Dleu	L-N-methylserine	Nmser
	D-lysine	Dlys	L-N-methylthreonine	Nmthr
25	D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
	D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
	D-phenylalanine	Dphe	L-N-methylvaline	Nmval
	D-proline	Dpro	L-N-methylethylglycine	Nmetg
	D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
30	D-threonine	Dthr	L-norleucine	Nle
	D-tryptophan	Dtrp	L-norvaline	Nva

WO 98/53061

	D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
	D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
	D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
	D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen
5	D-α-methylasparagine	Dmasn	α -methyl- α -napthylalanine	Manap
	D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
	D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D - α -methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
	D-α-methylleucine	Dmleu	α-napthylalanine	Anap
	D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
	D - α -methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D - α -methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D - α -methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D - α -methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	D-α-methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser
30	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp

	D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dommet
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
5	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
10	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	γ-aminobutyric acid	Gabu	N-(p-hydroxyphenyl)glycine	Nhtyr
	L-t-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L-α-methylalanine	Mala
15	L-α-methylarginine	Marg	L-α-methylasparagine	Masn
	L-α-methylaspartate	Masp	L-α-methyl-t-butylglycine	Mtbug
	L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
	L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
	L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
20	L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
	L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
	L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
	L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
25	L-α-methylserine	Mser	L-α-methylthreonine	Mthr
	L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr

WO 98/53061 PCT/AU98/00380

- 37 -

L-α-methylvaline Mval L-N-methylhomophenylalanine Nmhphe
N-(N-(2,2-diphenylethyl) Nnbhm N-(N-(3,3-diphenylpropyl) Nnbhe
carbamylmethyl)glycine carbamylmethyl)glycine
1-carboxy-1-(2,2-diphenyl- Nmbc
5 ethylamino)cyclopropane

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n=1 to n=6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N_α -methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and 15 the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

Such analogues also apply in respect of MCG4 and MCG7.

20

The present invention further contemplates chemical analogues of MCG18 capable of acting as antagonists or agonists of MCG18 or which can act as functional analogues of MCG18. Chemical analogues may not necessarily be derived from MCG18 but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to mimic certain physiochemical properties of MCG18. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening.

The identification of MCG:8 permits the generation of a range of therapeutic molecules capable of modulating expression of MCG18 or modulating the activity of MCG18. Modulators contemplated by the present invention includes agonists and antagonists of MCG18 expression. Antagonists of MCG18 expression include antisense molecules, ribozymes and co-suppression

molecules. Agonists include molecules which increase promoter ability or interfere with negative regulatory mechanisms. Agonists of MCG18 include molecules which overcome any negative regulatory mechanism. Antagonists of MCG18 include antibodies and inhibitor peptide fragments.

5

These types of modifications may be important to stabilise MCG18 if administered to an individual or for use as a diagnostic reagent.

Other derivatives contemplated by the present invention include a range of glycosylation variants 10 from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

Another embodiment of the present invention contemplates a method for modulating expression of MCG18 in a human, said method comprising contacting the mcg18 gene encoding MCG18 with an effective amount of a modulator of mcg18 expression for a time and under conditions sufficient to up-regulate or down-regulate or otherwise modulate expression of mcg18. For example, a nucleic acid molecule encoding MCG18 or a derivative thereof may be introduced into a cell to facilitate protection of that cell from becoming cancerous.

20

Another aspect of the present invention contemplates a method of modulating activity of MCG18 in a human, said method comprising administering to said mammal a modulating effective amount of a molecule for a time and under conditions sufficient to increase or decrease MCG18 activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative of MCG18 or a chemical analogue or truncation mutant of MCG18.

The present invention is further described with reference to the following non-limiting Examples.

EXAMPLE 1

A human gene (designated mcg4) was identified on chromosome 11q13 that on the basis of sequence homology is predicted to encode a putative transcription factor of 310 amino acids 5 (Fig. 1). mcg4 is transcribed in several different cell lines (Fig. 7).

EXAMPLE 2

The expressed sequence tag (EST) database contains partial sequence data for the murine (Fig. 10 2) and nematode (Fig. 3) homologues of mcg4.

EXAMPLE 3

MCG4 contains a sequence of cysteine residues within the N-terminal region of the protein that resembles zinc-finger binding domains of a novel type, ie. (HC₃)₂ [Fig. 4].

EXAMPLE 4

Sensitive sequence homology searches reveal that related cysteine-containing motifs are present in another *C. elegans* protein (Fig. 5) as well as the GATA-binding transcription factor from *S. pombe* (Fig. 6).

EXAMPLE 5

25 mcg4 will have commercial value due to its likelihood of encoding a novel transcription factor that is highly conserved amongst organisms, thus suggesting an integral role in gene regulation. mcg4 may also be involved in some way in tissue-specific or temporal regulation of certain genes, thus making it a potential target for modulating expression of those downstream effectors.

EXAMPLE 6

Nucleotide sequence data generated from cosmid clone cSRL-72c4 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) was aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul et al 1990) and was found to match numerous human and mouse entries (Table 4 and Figure 2). These matching ESTs were further used to identify overlapping entries in the EST database (Table 5). The nucleotide sequences of these human ESTs were complied using MacVector 4.2.1 software (IBI-Kodak) to produce the cDNA sequence shown in Figure 1. EST entries AA074703 and AA134788 are closely related at the nucleotide level to mcg4 and it is, therefore, likely that mcg4 is a member of a newly discovered gene family (Figure 8).

The cDNA sequence of mcg4 was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul et al, 1990) at 15 the National Center for Biotechnology Information (http://www.ncbi.nih.gov.nlm). As the protein appeared to be novel, a translation of the longest reading frame for the mcg4 cDNA was aligned to the EST database using the program TBLASTN, which performed a dynamic translation of the EST database in all 6 frames. The search results indicated that the nematode C. elegans had an MCG4-like protein (Figure 3), with the matching domains containing a spatial 20 sequence of Cysteine and Histidine residues which resembled a zinc-finger structure (Figure 4). The program BLASTP was used, therefore, to conduct sensitive searches of the protein databases for similar zinc-finger motifs. A weak match to the putative zinc-finger domain was observed for another protein from C. elegans (Figure 5) and a poorer match for the GATAbinding transcription factor from S. pombe (Figure 6). The putative initiation codon of human 25 mcg4 is not preceded by an in-frame stop codon and it is therefore possible that the cDNA described in Figure 1 is a truncated form. However, sequence alignment of human and mouse mcg4 ESTs showed a lower degree of nucleotide conservation prior to the assigned initiation codon, thus supporting the notion that the region represents the 5' UTR (Figure 9). To determine the expression pattern of mcg4, 15µg of the total cellular RNA (RNeasy Mini Kit, 30 Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% w/v MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer using 20 x SSC (Sambrook et al, 1989). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (32P-dCTP) cDNA probe (Church and Gilbert, 1984) for mcg4. After washes in 0.1 x SSC/0.1% w/v SDS at 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that mcg4 is expressed as a 1.6kb 5 message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 7).

EXAMPLE 7

A human gene (designated *mcg7*) was identified and isolated from chromosome 11q13 which 10 encodes a protein that bears striking homology with guanine nucleotide exchange factors (GEFs) from a wide variety of organisms (Fig. 12).

EXAMPLE 8

15 The composite mcg7 cDNA sequence is at least 2.4kb in length and Figure 13(a) shows a predicted translation product of at least 609 amino acids beginning at methionine 120. An alternative start site due to alternate exon splicing (indicated in lower case) may yield a protein of 671 amino acids starting at methionine 58 (Fig. 13a).

20 EXAMPLE 9

An mcg7 homologue from C. elegans has been identified, the product of which is highly conserved with that of MCG7 (Fig. 14). There are several salient features of the protein which have been underlined in Fig. 14 - namely: a guanine nucleotide binding region, a diacylglycerol binding region, and "EF-hand"-calcium binding regions. In addition, there are several potential cAMP, protein kinase C, and casein kinase II phosphorylation sites, as well as a number of potential sites for glycosylation (not indicated).

EXAMPLE 10

30

A number of partial human and murine EST clones exist for mcg7. The GenBank database

contains a cDNA (Acc. no. Y12336) encoding a full-length open reading frame (ORF) for human mcg7 as well as a partial murine mcg7 ORF (Y12339). In addition, the complete genomic sequence of the human mcg7 gene is contained within GenBank entry AC000134.

5 EXAMPLE 11

The best characterised GEFs are members of the family of ras oncoproteins, which play a pivotal role in signal transduction and when mutated are responsible for tumour development. A variety of therapeutic regimes for cancer treatment have been designed to specifically interfere with the ras signalling pathways. There is potential, therefore that the product of mcg7 could also be a target for such clinical strategies.

EXAMPLE 12

The nucleotide sequence for mcg7 cDNA was extended 5' with genomic DNA sequence from Genbank accession number AC000134 (positions 1-321) and analysed for additional coding sequence 5' to the putative initiation codon (nt 681-683) (Fig. 16). An additional in-frame ATG occurs at position nt 495-497 when the alternatively splice exon (position nt 504-609) is present (also shown in Fig. 13(a)). This closely matches the Kozak consensus. When this exon is absent, then the ATG is not in-frame and other possible initiation codons are absent (resulting translation shown in lower case lettering) (also shown in Fig. 13(b)). Further evidence that the initiation codon at position nt 681-683 is the true initiation site is given in Figure 15.

Alignment of human and a partial murine mcg7 cDNA sequences is shown in Figure 15. The putative initiation codon is at position nt 360-362. Both murine ESTs appear to have an upstream in-frame stop codon at position nt 326-328, downstream of the differentially spliced exon and the sequence alignment thus suggests that this region represents the 5' UTR of mcg7.

Furthermore, similarity with the *C. elegans* homologue strongly suggest that the ATG codon at position nt 360-362 encodes the N-terminus of MCG7.

EXAMPLE 13

Figure 17 shows data from experiments indicating that a truncated version of MCG7 when expressed as a GST fusion protein (construct B in Fig. 18) can function as a Ras-guanine nucleotide exchange factor. In brief, Ras (unprocessed and as a GST fusion protein) is loaded with ³H-GDP then incubated in the presence of excess cold GTP ± GST-MCG7. Full details of this assay can be found in Porfiri et al.

EXAMPLE 14

10

Nucleotide sequence data generated from cosmid clone cSRL-20h12 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) were aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul et al, 1990) and was found to match GenBank entries T78563 (clone 113434) TO9103 (clone HIBBP12) and AA035643 (clone 471819). EST clones 113434 and 471819 were obtained from Genome Systems Inc. and these DNAs were sequenced on both strands with gene-specific primers (Table 5) to generate the cDNA sequence of mcg7 shown in Figures 13(a) and (b).

The cDNA sequence of *mcg*7 was translated in all possible reading frames and compared to the 20 GenBank non-redundant protein database using the program BLASTX (Altschul *et al*, 1990) and the coding region was assigned on the basis of showing homology to the *C. elegans* protein F25B3.3 (Figure 14). The *mcg*7 cDNA composite was suspected to contain a single nucleotide error that originated from clone 471819 and the correct nucleotide sequence was, therefore, sought by reverse transcription-polymerase chain reaction (RT-PCR) of the cDNA fragment from a human cDNA pool. Total RNA was extracted from a human lymphoblastoid cell line using an RNeasy Mini Kit (Qiagen). cDNA synthesis was conducted with the reverse transcriptase Superscript II RNaseH- (GIBCO, BRL) and random hexamers using the procedure recommended by the manufacturer (GIBCO, BRL). One fortieth of the cDNA mix was subjected to 35 cycles of PCR using the following cycling conditions: 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 90 seconds. The 50µl reaction mix consisted of 1x reaction buffer (Dade Scientific), 2mM dNTP mix, 20pmol of primers (see Table 6) MCG7UF (within the

variably spliced exon of Figure 13(b), between nucleotide positions 184-201) and SGCADRV2 (between nucleotide positions 866-846 of Figure 13(a)) and 10 units of Dynazyme (Dade Scientific). The resulting PCR product was cloned into the pGEM-T vector (Promega) using standard methodology and sequenced using gene-specific primers. The correct nucleotide sequence of mcg7 (as shown in Figure 13(a)) matches that of the recently release GenBank entry Y12336. A partial mouse mcg7 cDNA sequence can also be found in GenBank entry Y12339.

EXAMPLE 15

10 The coding sequence of mcg7 was cloned into vectors for expression in both bacterial and mammalian cells. In addition to the full-length constructs, the deletion constructs shown in Figure 18 were designed to retain the guanine nucleotide exchange (GEF) domain. For prokaryotic expression, the mcg7 coding region was inserted downstream of and in-frame with the Sj26 cassette of the pGEX (Pharmacia) series of vectors (Smith and Johnson, 1988) using standard cloning techniques (Sambrook et al, 1989). For mammalian expression, the mcg7 coding sequence was first myc-tagged at the N-terminus and then ligated into the expression vector pc Exv-n using standard cloning techniques. Ligation junctions of the constructs were sequences as the cloning strategies inadvertently changed or introduced additional amino acids as shown below.

20

Construct (A): EST clone 113434 was digested with ApaI (Figure 13(a), nucleotide positions 1022 to >2416 (within the vector)), blunt-ended with T4 DNA polymerase according to the specifications of the manufacturer (New England Biolab) and ligated into the SmaI site of pGEX-3X.

25

Sequence of the pGEX and mcg7 (underlined) junction:

đ

pGEX-3X mcg7 (1022)

Sj26 ... GGG ATC CCC CTG GTC [SEQ ID NO:19]

additional amino acids Gly Ile Pro

30

Construct (B): EST clone 113434 was digested with EcoRI (Figure 13(a), nucleotide

positions <695 (within the vector) to 1711) and ligated into the EcoRI site of pGEX-1.

Sequence of the pGEX and mcg7 (underlined) junction:

pGEX-1

mcg7 (695)

5 Sj26 ... GAA TTC GGC ACG AG<u>C CGA CGG</u> [SEQ ID NO:20] additional amino acids Glu Phe Gly Thr Ser

Construct (C): full-length mcg7: The pGEM-T clone containing the 5' end of the mcg7 coding region was digested with ApaI (subsequently blunt-ended with T4 DNA polymerase) and BstXI to liberate the fragment between nucleotide positions 336 and 830 of Figure 13(a). Clone 113434 was digested with BstXI and HindIII (vector derived) to liberate a fragment between nucleotide positions 830 > and 2416 (vector derived) of Figure 13(a). A pGEM-11zf vector (Promega) containing the myc-tag was digested with ApaI (subsequently blunt-ended with T4 DNA polymerase) and HindIII, and ligated with the 2 inserts described above.

15

Sequence of the myc-tag/mcg7 junction [SEQ ID NOs:21/22]:

ATGGAGCAGAAGCTGATCTCCGAGGAGGACCTG CCCGGGGCAGCTggatccG CAGCCCACCCCGCGCCGCGCCATG

20 M E Q K L I S E E D L P G A A G S A A H P A P A A M

------additional amino acids-----

The myc-tagged full-length mcg7 insert in pGEM-11zf was then excised with SacI and HindIII (both vector derived) and directionally cloned into the mammalian expression vector pEXV 25 (Beranger et al, 1994).

Construct (D): Construct (C) in pGEM-11zf was sequentially digested with *Hind*III (this site was subsequently blunt-ended with T4 DNA polymerase) then *BamH*I, and ligated into pGEX-2T digested with *BamH*I and *SmaI*. Digestion with *BamH*I, and ligated into pGEX-2T digested with *BamH*I and *SmaI*. Digestion with *BamH*I removed the *myc*-tag of Construct (C).

Sequence of the pGEX and mcg7 [SEQ ID NO:23/24] (underlined) junction:

- 46 -

-----additional amino acids------

pGEX-2 BamHI mcg7 (337)

Sj26 ... gga tcc GCA GCC CAC CCC GCG CCG GCC ATG

Gly Ser Ala Ala His Pro Ala Pro Ala Met

5

EXAMPLE 16

Overnight bacterial cultures containing the pGEX plasmid were used to inoculate 500ml of Luria Broth media containing $50\mu g/ml$ ampicillin. The cultures were grown to an OD of ~0.8 and then 10 induced with 1mM of IPTG for up to 3 hours at 37°C. The bacteria were pelleted and resuspended in 15 ml of STE buffer (10mM Tris pH 8.0, 150 mM NaCl and 1mM EDTA) with 1 mg/ml lysozyme. The mixture was left on ice for more than 1 hour and subsequent steps were performed at 4°C. Protease inhibitors aprotinin, pepstatin and leupeptin were added at final concentrations of 25µg/ml, prior to the addition of Triton-X-100 (2% v/v final) and n-lauroyl 15 sarcosine (1.5% w/v final). The lysate was sonicated for ~1 minute and pelleted at 14,000 x g for 15 minutes. 100 μ l of 50% w/v glutathione-sephadex bead slurry (in PBS) was added per ml of supernatant. Following a 30 minute incubation at 4°C, the beads were washed three times with NETN (20mM Tris-HC1 pH 8.0, 100mM NaCl, 1mM EDTA, 0.5% NP40), once with NETN-HS (equivalent to NETN but with 1M NaCl), and once in NETN. The bound protein 20 was directly analysed by SDS-polyacrylamide gel electrophoresis (PAGE) as described below or the bound protein was eluted from the beads with the following elution buffer (50mM Tris pH 8.0, 150mM NaCl, 5mM MgCl₂, 1mM DTT, 10mM reduced glutathione) for use in GDP release assays.

25

EXAMPLE 17

Twenty microlitres of GST-sepharose-bound MCG7 were added to an equal volume of 2 x 30 sample loading dye (100mM Tris pH6.8, 2% v/v mercaptoethanol, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol), boiled for 5 min and loaded onto a 7.5% w/v SDS-PAGE gel (Sambrook *et al*, 1989). The Coomassie brilliant blue stained gel (Sambrook *et al*, 1989)

typically displayed a protein doublet, running between 87-95 kDa consisting of the MCG7-GST fusion and a slightly smaller, co-purified contaminating *E. coli* protein of ~105kDa. The calculated molecular weight of full-length MCG7 is 77.5 kDa (Construct (D)) and the GST component has a molecular weight of 26kDa, hence, the recombinant protein runs slightly smaller than predicted. A Western blot of the same gel probed with anti-GST antibody yields an MCG7-specific band at the same position as that of the stained gel.

EXAMPLE 18

10 Assumptions: (a) GST-Ras molecular weight = 50 kD; (b) Concentration of GST-Ras solution = 1mg/ml = 20μM; (c) [³H]-GDP is 1mCi/ml and 13.3Ci/mmol, therefore [H]-GDP concentration = 75 μM and 1pmol [³H]-GDP=15,466 cpm; (d) Elution buffer = Buffer E = 20 mM Tris-Cl, pH7.5; 50mM NaCl; 5mM MgCl₂; 1mM DTT (added just before use). Buffer E + BSA= Buffer E+1mg/ml BSA (added just before use).

15

Mix together, in the following order and mix well after each addition:

10μl (=10μg) GST-Ras (@1mg/ml in Buffer E), 463μl Buffer E + BSA, 7μl [³H]-GDP, 10ml 490 μM EDTA. Incubate @ RT for 10 min. Add 10μl 0.5 M MgCl₂ and mix well. Incubate @ RT for 10 min. Place on ice. During the first incubation the excess EDTA concentration is 20 5mM, during the second incubation the excess Mg concentration is 5mM. The [³H]-GDP concentration is 1μM and the final concentration of GST-Ras is 400nM. Thus 20ml of the final mix will contain 8pmol of GST-Ras protein. Specific activity of GDP is 15,446 cpm/pmol x (1/1.4) = 11,047 cpm/pmol.

25

EXAMPLE 19

Exchange Ras with labelled GDP as above. Add unlabelled GTP (stock = 100mM, pH7) to 1 mM. Adjust Mg concentration by adding 5μl 0.5 EDTA to labelled Ras, 5μl 0.5M EDTA to 500μl MCG7, and 5μl 0.5M EDTA to 500μl Buffer E + BSA. On ice set up microfuge tubes with 40μl Ras-GDP (in triplicate) with 40μl MCG7 or Buffer E + BSA (control). Transfer tubes to heat block @ 25°C and incubate for 10, 20 or 30 min. Stop exchange reactions with 1ml of

ice cold buffer E and place on ice. Pre-soak nitrocellulose filters, pore size 45μm, in Buffer E. Assemble the vacuum manifold apparatus (Millipore) with wet filters and plug the wells with rubber bunds. Switch on the vacuum pump. Remove the first plug, aliquot the sample and once it has been sucked through, wash the filter with 10ml of ice cold Buffer E. Remove next plug etc and continue round the manifold. Take manifold apart. Pin the filters to a pin board reserved for [³H]. Air dry. Take up in 4ml scintillation fluid and count. These studies have been carried out with a truncated MCG7-GST fusion protein (amino acids 341 of Figure 13a to stop encoded within construct B).

10

EXAMPLE 20

A human gene was identified from chromosome 11q13 that encodes a new member of the DnaJ family of proteins (designated MCG18). This gene (mcg18) is expressed as an ~1.4kb mRNA (Fig. 28) and is predicted to encode a 241 amino acid product (Fig. 19).

15

EXAMPLE 21

MCG18 has partial homology to E. coli dnaJ and other human DnaJ family members in that it contains the J domain (Fig. 20).

20

EXAMPLE 22

MCG18 has greatest homology to functionally undefined proteins from *C. elegans* (Fig. 21) and *S. pombe* (Fig. 22) that also feature the J domain but maintain sequence similarity through the central and C-terminal regions of the proteins.

EXAMPLE 23

The J domain is proposed to mediate interaction with heat shock protein (Hsp70) 70 and consist 30 of some 70 amino acids, frequently located at the N-terminus of the protein. One of these proteins, tumorous imaginal discs (Tid58) from *Drosophila virilis* (Fig. 23) functions as a

tumour suppressor.

EXAMPLE 24

5 A comparison of homology between MCG18 and human DnaJ proteins HDJ-2/H5DJ, HDJ-1/HSP40 and HSJ1 is shown in Fig. 24.

EXAMPLE 25

- 10 During the sequence characterisation of the VRF/VEGFB promoter region on cosmid CLGW4
 [Grimmond et al, 1996], which maps to chromosome 11q13 the inventors identified a sequence that exactly matched numerous human and mouse expressed sequence tags (ESTs) in the EST database from a gene which we designated mcg18. EST clones for human (GenBank accession number T69741, clone 108172; accession number H40901, clone 177008) and mouse mcg18
 15 (accession number W34884, clone 350966; accession number W64183, clone 385535) were obtained from Genome Systems Inc. and sequenced with the gene-specific primers shown in Table 7. The EST clones listed in Table 8 were also utilised in generating the full-length coding sequence for human (Figure 19) and mouse (Figure 25) mcg18. The EST database also contained mcg18 cDNA entries that were alternately (or partially) spliced, and in order to understand their ability to encode new polypeptides, the gene structure of mcg18 was determined by sequencing human and mouse genomic templates with gene-specific primers.
- Genomic fragments containing the human [Grimmond et al, 1996] and murine genes [Townson et al, 1996] have been previously reported. Cosmid CLGW4 contains the entire human gene and λ121 contains the entire mouse gene, as determined by direct sequencing of the templates with the oligonucleotides listed in Table 7. Plasmids containing sub-fragments of λ121 and cosmid CLGW4 were prepared using plasmid purification kits (Qiagen) and sequenced as described previously [Grimmond et al, 1996; Townson et al, 1996] using primers designed against cDNA and genomic sequences. The BLAST suite of programs [Altschul et al, 1990] was used to compare the sequence data against the nucleotide and protein databases at the National Center for Biotechnology Information (http//www.ncbi.nih.gov.nlm). The sequence

data were compiled using MacVector 4.2.1 software (IBI-Kodak). ClustalW sequence alignments [Thompson *et al*, 1994] were conducted using the Australian National Genome Information Service computer faculty at the University of Sydney, Australia.

- 5 The cDNA sequence of human mcg18 (Figure 19) was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX [Altschul et al, 1990] and the coding region was identified on the basis of showing homology to the DnaJ family of proteins (Figure 20). The DnaJ domain is encoded within the longest open reading frame and the assigned initiation codon is preceded by an in-frame stop codon (Figure 27). Similar database search results were obtained for the mouse mcg18 cDNA, and the alignment of human and mouse protein sequences is shown in Figure 26. MCG18 has greatest homology to gene products from C. elegans (Figure 21) and S. pombe (Figure 22). Although it shares a similar J-domain, MCG18 does not contain other domains described for the tumour suppressor gene from D. virilis (Figure 23), nor is it a homologue of other reported human J-domain-containing proteins (Figure 24).
- To determine the expression pattern of mcg18, $15\mu g$ of total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer using 20 x SSC (Sambrook et al, 1986). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (32 P-dCTP) cDNA probe (Church and Gilbert, 1984) for mcg18. After washes in 0.1 x SSC/0.1% w/v SDS for 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that mcg18 is expressed as a 1.4kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 28).

- 51 -

TABLE 4

ESTs matching mcg4

accession number	seq. run	organism	score	e E value	N
gb AA399110 AA39911) zt89e06.sl	Soares testis NHT Homo sa	1136	4.0e-168	2
ap n39615 n39615	yy51g06.s1	Homo sapiens cDNA clone 2	1521	5.3e-168	4
gb AA514406 AA51440	n£57d01.s1	NCI_CGAP_Co3 Homo sapiens	931	5.5e-166	3
gb AA544946 AA544946	vk38e02.rl	Soares mouse mammary glan	1207	8.4e-164	2
gb AA450076 AA450076		Soares total fetus Nb2HF8	691	2.3e-160	4
gb AA535731 AA535731		NCI_CGAP_Col Homo sapiens	796	3.5e-158	4
gb w79710 w79710	zd86f01.rl	Soares fetal heart NbHH19	1644	1.1e-157	4
gb AA503531 AA503531	ne47e08.s1	NCI_CGAP_Col Homo sapiens	736	4.0e-156	4
gb AA450132 AA450132		Soares total fetus Nb2HF8	1955	3.9e-155	1
ap vy38068 vy38068		Soares testis NHT Homo sa	1315	5.4e-148	2
gb W60405 W60405		Soares fetal heart NbHH19	1022	1.8e-139	4
gb W81382 W81382	zd86f01.sl	Soares fetal heart NbHH19	605	3.5e-125	5
gb AA047617 AA047617		Soares fetal heart NbHH19	922	4.6e-125	2
gb AA282175 AA282175		NCI_CGAP_GCB1 Homo sapien	1577	2.0e-123	1
gb AA242159 AA242159		Barstead mouse pooled org	866	7.7e-117	2
gb AA068680 AA068680		Stratagene mouse embryoni	1280	1.6e-98	1
gb W46766 W46766	zc36b07.s1	Soares senescent fibrobla	506	9.6e-92	3
gb N93704 N93704	zb51c04.s1	Soares fetal lung NbHL19W	584	9.0e-91	4
gb AA155210 AA155210	mr98e01.rl	Stratagene mouse embryoni	840	7.6e-87	2
gb AA366022 AA366022	EST76915 Pi	neal gland II Homo sapien	1077	2.4e-81	1
gb AA037691 AA037691	zk34h12.s1	Soares pregnant uterus Nb	949	2.1e-80	2
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor	1016	3.1e-76	1
dbj C00696 C00696	HUMGS000825	1, Human Gene Signature,	1009	1.2e-75	1
gb T98249 T98249	ye59a07.sl	Homo sapiens cDNA clone 1	998	6.7e-75	1
gb W21588 W21588		Soares fetal lung NbHL19W	484	1.1e-69	4
gb H32171 H32171	EST107015 R	attus sp. cDNA 5' end.	828	1.1e-60	1
gb AA108092 AA108092	nm89e06.rl	Stratagene mouse embryoni	782	1.3e-60	2
gb AA017857 AA017857	mh44d10.r1	Soares mouse placenta 4Nb	665	2.5e-60	2
gb AA037690 AA037690	zk34h12.r1	Soares pregmant uterus Nb	540	9.4e-53	2
gb AA531006 AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapien	535	5.4e-48	2
gb N46760 N46760	yy51g06.rl	Homo sapiens cDNA clone 2	665	9.5e-47	1
gb W23584 W23584		Soares fetal heart NoHH19	457	1.8e-44	2
gb W42214 W42214	mc69h09.rl	Soares mouse embryo NbMEl	460	1.3e-38	3
gb AA244877 AA244877		Soares mouse NML Mus musc	429	2.9e-25	1
gb W32939 W32939	zc07h03.rl	Soares parathyroid tumor	320	4.8e-18	1

- 52 -

TABLE 5
ESTs matching AA074703 (mcg4-related cDNA)

Database: Non-redundant Database of GenBank EST Division 1,222,625 sequences; 449,352,662 total letters.

Smallest

Sum

				Sum	
			High	Probabil:	ity
Sequences producing H	ligh-scoring	Segment Pairs:	Score	P (N)	N
accession number	seq. run	organism	score	E value	N
gb AA074703 AA074703	zm76g07.rl	Stratagene neuroepitheli	2071	4.0e-167	1
gb AA068680 AA068680	mm61a05.rl	Stratagene mouse embryon	1270	4.4e-145	4
gb AA134788 AA134788	zm81g02.r1	Stratagene neuroepitheli	946	1.3e-144	5
gb AA399110 AA399110	zt89e06.s1	Soares testis NHT Homo s	520	8.7e-119	6
gb N39612 N39612	yy51g06.s1	Homo sapiens cDNA clone	582	9.6e-110	7
gb AA282175 AA282175	zt02d03.s1	NCI_CGAP_GCB1 Homo sapie	771	9.4e-80	3
gb W81382 W81382	zd86f01.s1	Soares fetal heart NbHH1	329	1.6e-75	6
gb AA544946 AA544946	vk38e02.r1	Soares mouse mammary gla	644	9.6e-63	2
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor	294	4.5e-42	4
gb W57106 W57106	md57c12.r1	Soares mouse embryo NbME	394	1.9e-30	2
gb AA244877 AA244877	mx25a04.rl	Soares mouse NML Mus mus	162	2.1e-27	4
gb AA017857 AA017857	mh44d10.r1	Soares mouse placenta 4N	230	3.7e-23	3
gb AA531006 AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapie	139	2.3e-19	3
gb H32171 H32171	EST107015 R	attus sp. cDNA 5' end.	207	2.6e-10	2
gb W79710 W79710	zd86f01.r1	Soares fetal heart NoHH1	157	0.0073	1

TABLE 6
mcg7-specific oligonucleotides

name	sequence (5' to 3')	SEQ ID NOs.
M1044R	GGA CAA AGT GTG TGA TGA ACC	SEQ ID NO:25
MCG7-GEF-REV2	CTC ATC CTC CGT CTG ATA CTG	SEQ ID NO:26
M7R	GTA GAT GTG GAT CAG CTT GG	SEQ ID NO:27
MCG7 CA FOR	AGG TGG AGA ATG GTC AAGG	SEQ ID NO:28
MCG7-GEF-REV	GTC ATA GTC TGT CTC CTA CT	SEQ ID NO:29
MCG7 GEF FOR	ACA TAG ACA GCG TGC CTA CC	SEQ ID NO:30
MCG7-PKC-REV	TAC AAC CTT AGG GAC ACC AG	SEQ ID NO:31
MCG7-PKC-FOR	TGC TGA GCC TGC TCA CGG TG	SEQ ID NO:32
T09103F	CAA GTG AAC AGC ACG TCC	SEQ ID NO:33
M7F	GAC TAT CTC AAG GAC CAG CTG	SEQ ID NO:34
MCG7UF	GGT TCG GTC CGA GCC CGG	SEQ ID NO:35
SGCADRV2	GGA GCG ATA CTC CAA GTA GGT	SEQ ID NO:36
	M1044R MCG7-GEF-REV2 M7R MCG7 CA FOR MCG7-GEF-REV MCG7 GEF FOR MCG7-PKC-REV MCG7-PKC-FOR T09103F M7F MCG7UF	M1044R GGA CAA AGT GTG TGA TGA ACC MCG7-GEF-REV2 CTC ATC CTC CGT CTG ATA CTG M7R GTA GAT GTG GAT CAG CTT GG MCG7 CA FOR AGG TGG AGA ATG GTC AAGG MCG7-GEF-REV GTC ATA GTC TGT CTC CTA CT MCG7 GEF FOR ACA TAG ACA GCG TGC CTA CC MCG7-PKC-REV TAC AAC CTT AGG GAC ACC AG MCG7-PKC-FOR TGC TGA GCC TGC TCA CGG TG T09103F CAA GTG AAC AGC ACG TCC M7F GAC TAT CTC AAG GAC CAG CTG MCG7UF GGT TCG GTC CGA GCC CGG

TABLE 7

mcg18-SPECIFIC OLIGONUCLEOTIDES

name	sequence 5' to 3'
5 HVESTF	AGC GGG CCA GGC CCC TTC [SEQ ID NO:37]
HV195F	CAT CCT GGT CCA ATG CGC TC [SEQ ID NO:38]
HV387F2	GCA CTG AGG AAG TTA AAC GAG C [SEQ ID NO:39]
HV408R	GCT CGT TTA ACT TCC TCA GTG C [SEQ ID NO:40]
EXON1REV	GCT CAG CTC CAC AAA GCG GCT [SEQ ID NO:41]
0 HVEST426F	ACC AGC TCC GCT CAG GTA G [SEQ ID NO:42]
HVEST623R	TCC AGG AGC TGT GTG TTT GG [SEQ ID NO:43]
SGVESTF3	CCA GTT TCA CAG CGT GAG G [SEQ ID NO:44]
HVEST631R	CAG CAT GAG GAG GCA G [SEQ ID NO:45]

TABLE 8 EST CLONE SEQUENCES USED TO GENERATE HUMAN AND MOUSE mcg18~cDNA~SEQUENCE~COMPOSITES

EST clone number	organism	GenBank accession number
lg2815	human	D45683
001-T2-18	human	F17225
273748	human	N37043
177008	human	H40901 and H40939
258011	human	N30776
276887 -	human	N44004
108172	human	T69741
307529	human	W21083 and W32579
342027	human	W60283
354288	mouse	W44038
350966	mouse	W348844
426261	mouse	AA002868
368185	mouse	W53911
385535	mouse	W64183
404472	mouse	W82959
406437	mouse	W83482

BIBLIOGRAPHY

- 1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) J. Mol. Biol. 215: 403-410.
- 2. Church, G., and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 18: 1991-1995.
- 3. Porfiri et al. J. Biol. Chem. 269: 22672-22677 (1994).
- 4. Sambrook, J., Frtisch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbour Laboratory, Cold Spring Harbour, NY, USA.
- 5. Smith, D.B., and Johnson, K.S. (1988) Gene 67: 31-40.
- 6. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) Nucleic Acids Res. 22: 4673-4680.
- 7. Beranger, F., Paterson, H., Powers, S., de Gunzburg, J. and Hancock, J.F. (1994)

 Molecular and Cellular Biology 14: 744-758.
- 8. Grimmond, S., Lagercrantz, J., Drinkwater, C., Silins, G., Townson, S., Pollock, P., Gotley, D., Carson, E., Rakar, S., Nordenskjöld, M., Ward, L., Hayward, N., and Weber, G (1996) *Genome Res. 6:* 124-131.
- 9. Townson, S., Lagercrantz, J., Grimmond, S., Silins, G., Nordenskjöld, Weber, G., and Hayward, N. (1996) *Biochem. Biophys. Res. Commun.* 220: 922-928.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: (OTHER THAN US):

The Council of The Queensland Institute of

Medical Research

(US ONLY): HAYWARD Nicholas, SILINS Ginters, GRIMMOND Sean, GARTSIDE Michael and HANCOCK, John

- (ii) TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
- (iii) NUMBER OF SEQUENCES: 45
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: DAVIES COLLISON CAVE
 - (B) STREET: 1 LITTLE COLLINS STREET
 - (C) CITY: MELBOURNE
 - (D) STATE: VICTORIA
 - (E) COUNTRY: AUSTRALIA
 - (F) ZIP: 3000
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT INTERNATIONAL
 - (B) FILING DATE: 22-MAY-1998
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PO6973
 - (B) FILING DATE: 23-MAY-1997
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PO6974
 - (B) FILING DATE: 23-MAY-1997
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PO6972
 - (B) FILING DATE: 23-MAY-1997

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1459
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1460
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1458
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: HUGHES, DR E JOHN L
- (C) REFERENCE/DOCKET NUMBER: EJH/AF

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: +61 3 9254 2777
- (B) TELEFAX: +61 3 9254 2770
- (C) TELEX: AA 31787

120

(2)	INFORMATION	FOR	SEQ	ID	NO:1
-----	-------------	-----	-----	----	------

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids(B) TYPE: amino acid(C) STRANDEDNESS: single

 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Cys Xaa Xaa Cys Xaa Gly Xaa Gly

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1242 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (ix) FEATURE:

105

- (A) NAME/KEY: CDS
 (B) LOCATION: 30..959
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

110

	TCAC	gta'a <i>i</i>	ACA (CAGA	GACTY	GG G(GATC	GATC							CCC Pro		53
,	AGA Arg	AAG Lys 10	GTG Val	ACC Thr	AAC Asn	CTG Leu	TTC Phe 15	TGC Cys	TTC Phe	GAA Glu	CAT His	CGG Arg 20	GTC Val	AAC Asn	GTC Val	TGC Cys	101
	GAG Glu 25	CAC His	TGC Cys	CTG Leu	GTA Val	GCC Ala 30	AAT Asn	CAC His	GCC Ala	AAG Lys	TGC Cys 35	ATC Ile	GTC Val	CAG Gln	TCC Ser	TAC Tyr 40	149
	CTG Leu	CAA Gln	TGG Trp	CTC Leu	CAA Gln 45	GAT Asp	AGC Ser	GAC Asp	TAC Tyr	AAC Asn 50	CCC Pro	AAT Asn	TGC Cys	CGC Arg	CTG Leu 55	TGC Cys	197
	AAC Asn	ATA Ile	CCC Pro	CTG Leu 60	GCC Ala	AGC Ser	CGA Arg	GAG Glu	ACG Thr 65	ACC Thr	CGC Arg	CTT Leu	GTC Val	TGC Cys 70	TAT Tyr	GAT Asp	245
	CTC Leu	TTT Phe	CAC His 75	TGG Trp	GCC Ala	TGC Cys	CTC Leu	AAT Asn 80	GAA Glu	CGT Arg	GCT Ala	GCC Ala	CAG Gln 85	CTA Leu	CCC Pro	CGA Arg	293
	AAC Asn	ACG Thr 90	GCA Ala	CCT Pro	GCC Ala	GGC Gly	TAT Tyr 95	CAG Gln	TGC Cys	CCC Pro	AGC Ser	TGC Cys 100	AAT Asn	GGC Gly	CCC Pro	ATC Ile	341
	TTC Phe	Pro	CCA Pro	ACC Thr	AAC Asn	CTG Leu	Ala	GGC Gly	CCC Pro	Val	GCC Ala	Ser	GCA Ala	CTG Leu	AGA Arg	GAG Glu	389

	CTG Leu															437
ATC Ile	GAT Asp	GAG Glu	GTG Val 140	GTG Val	AGC Ser	CCA Pro	GAG Glu	CCC Pro 145	GAG Glu	CCC Pro	CTC Leu	AAC Asn	ACG Thr 150	TCT Ser	GAC Asp	485
	TCT Ser															533
GAG Glu	GTA Val 170	GAC Asp	AGC Ser	GCC Ala	TCT Ser	GCT Ala 175	GCC Ala	CCA Pro	GCC Ala	TTC Phe	TAC Tyr 180	AGC Ser	CGA Arg	GCC Ala	CCC Pro	581
	CCC Pro															629
	GGC Gly															677
	CGG Arg															725
	AAG Lys															773
	AGC Ser 250															821
	GGG Gly															869
	CTC Leu														AAC Asn	917
	GAC Asp													TGA *		962
GCC	CCT	rgc :	rtgt	GCT	AG G	CCAG	CCTAC	G GA	rgtgo	GTT	CTG	rgga	GGA (GAGG	CGGGGT	1022
AATO	GGGG2	AGG (TGA	GGC2	AC C	rctt	CACTO	G CC	CTC	rccc	TCA	AGCC	TAA (GACA	CTAAGA	1082
CCC	CAGAC	CCC 2	AAAG	CCAA	T C	CACC	AGAG'	r gg	CTCG	CAGG	CCA	GGCC'	rgg i	AGTC	CCCGTG	1142
GGT	CAAGO	CAT	r tgt (CTTG	AC T	rgct'	TTCT	c cc	GGT	CTCC	AGC	CTCC	GAC (CCT	cgcccc	1202
ATG	AAGG2	AGC 1	rggcz	AGGT	GG A	ATAA	AACA	A CA	ACTT!	TATT						1242

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 310 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly Leu Cys Lys Cys Pro Lys Arg Lys Val Thr Asn Leu Phe Cys Phe Glu His Arg Val Asn Val Cys Glu His Cys Leu Val Ala Asn His Ala Lys Cys Ile Val Gln Ser Tyr Leu Gln Trp Leu Gln Asp Ser Asp 35 40 45 Tyr Asn Pro Asn Cys Arg Leu Cys Asn Ile Pro Leu Ala Ser Arg Glu Thr Thr Arg Leu Val Cys Tyr Asp Leu Phe His Trp Ala Cys Leu Asn 65 70 75. Glu Arg Ala Ala Gln Leu Pro Arg Asn Thr Ala Pro Ala Gly Tyr Gln Cys Pro Ser Cys Asn Gly Pro Ile Phe Pro Pro Thr Asn Leu Ala Gly 105 Pro Val Ala Ser Ala Leu Arg Glu Lys Leu Ala Thr Val Asn Trp Ala Arg Ala Gly Leu Gly Leu Pro Leu Ile Asp Glu Val Val Ser Pro Glu Pro Glu Pro Leu Asn Thr Ser Asp Phe Ser Asp Trp Ser Ser Phe Asn Ala Ser Ser Thr Pro Gly Pro Glu Glu Val Asp Ser Ala Ser Ala Ala Pro Ala Phe Tyr Ser Arg Ala Pro Arg Pro Pro Ala Ser Pro Gly Arg Pro Glu Gln His Thr Val Ile His Met Gly Asn Pro Glu Pro Leu Thr His Ala Pro Arg Lys Val Tyr Asp Thr Arg Asp Asp Asp Arg Thr Pro Gly Leu His Gly Asp Cys Asp Asp Lys Tyr Arg Arg Arg Pro Ala 225 230 235 Leu Gly Trp Leu Ala Arg Leu Leu Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Asp Ser Asp Pro Asn Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH: 2415 base pairs
(B)	TYPE: nucleic acid
(C)	STRANDEDNESS: single
(D)	TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 3..2188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CG AT	_				_		TCC (Ser)						47
CAT (95
CCA C							Cys						143
AGA (191
TGT (Glu			239
CAG (Gln (80													287
GGG (-	_										335
CCA (Ala						383
AAG (Leu							431
GAT (Arg			479
ATG A Met 1 160									Leu				527
CTC (Leu)								Asn				Gln	575
GTG . Val			His				Trp				Pro		623

GAG Glu	TTT Phe	GAC Asp 210	TTG Leu	AAC Asn	CCG Pro	GAG Glu	TTG Leu 215	GCT Ala	GAG Glu	CAG Gln	ATC Ile	AAG Lys 220	GAG Glu	CTG Leu	AAG Lys	671
GCT Ala	CTG Leu 225	CTA Leu	GAC Asp	CAA Gln	GAA Glu	GGG Gly 230	AAC Asn	CGA Arg	CGG Arg	CAC His	AGC Ser 235	AGC Ser	CTA Leu	ATC Ile	GAC Asp	719
ATA Ile 240	GAC Asp	AGC Ser	GTC Val	CCT Pro	ACC Thr 245	TAC Tyr	AAG Lys	TGG Trp	AAG Lys	CGG Arg 250	CAG Gln	GTG Val	ACT Thr	CAG Gln	CGG Arg 255	767
AAC Asn	CCT Pro	GTG Val	GGA Gly	CAG Gln 260	AAA Lys	AAG Lys	CGC Arg	AAG Lys	ATG Met 265	TCC Ser	CTG Leu	TTG Leu	TTT Phe	GAC Asp 270	CAC His	815
CTG Leu	GAG Glu	CCC Pro	ATG Met 275	GAG Glu	CTG Leu	GCG Ala	GAG Glu	CAT His 280	CTC Leu	ACC Thr	TAC Tyr	TTG Leu	GAG Glu 285	TAT Tyr	CGC Arg	863
TCC Ser	TTC Phe	TGC Cys 290	Lys	ATC Ile	CTG Leu	TTT Phe	CAG Gln 295	GAC Asp	TAT Tyr	CAC His	AGT Ser	TTC Phe 300	GTG Val	ACT Thr	CAT His	911
GGC Gly	TGC Cys 305	ACT Thr	GTG Val	GAC Asp	AAC Asn	CCC Pro 310	GTC Val	CTG Leu	GAG Glu	CGG Arg	TTC Phe 315	ATC Ile	TCC Ser	CTC Leu	TTC Phe	959
AAC Asn 320	AGC Ser	GTC Val	TCA Ser	CAG Gln	TGG Trp 325	GTG Val	CAG Gln	CTC Leu	ATG Met	ATC Ile 330	CTC Leu	AGC Ser	AAA Lys	CCC Pro	ACA Thr 335	1007
GCC Ala	CCG Pro	CAG Gln	CGG Arg	GCC Ala 340	CTG Leu	GTC Val	ATC Ile	ACA Thr	CAC His 345	TTT Phe	GTC Val	CAC His	GTG Val	GCG Ala 350	Glu	1055
AAG Lys	CTG Leu	CTA Leu	CAG Gln 355	CTG Leu	CAG Gln	AAC Asn	TTC Phe	AAC Asn 360	ACG Thr	CTG Leu	ATG Met	GCA Ala	GTG Val 365	GTC Val	GGG Gly	. 1103
GGC Gly	Leu	AGC Ser 370	CAC His	AGC Ser	TCC Ser	ATC Ile	TCC Ser 375	CGC Arg	CTC Leu	AAG Lys	GAG Glu	ACC Thr 380	CAC His	AGC Ser	CAC His	1151
GTT Val	AGC Ser 385	CCT Pro	GAG Glu	ACC Thr	ATC Ile	AAG Lys 390	CTC Leu	TGG Trp	GAG Glu	GGT Gly	CTC Leu 395	ACG Thr	GAA Glu	CTA Leu	GTG Val	1199
ACG Thr 400	GCG Ala	ACA Thr	GGC Gly	AAC Asn	TAT Tyr 405	GGC Gly	AAC Asn	TAC Tyr	CGG Arg	CGT Arg 410	CGG Arg	CTG Leu	GCA Ala	GCC Ala	TGT Cys 415	1247
GTG Val	GGC Gly	TTC Phe	CGC Arg	TTC Phe 420	CCG Pro	ATC Ile	CTG Leu	GGT Gly	GTG Val 425	CAC His	CTC Leu	AAG Lys	GAC Asp	CTG Leu 430	GTG Val	1295
GCC Ala	CTG Leu	CAG Gln	CTG Leu 435	GCA Ala	CTG Leu	CCT Pro	GAC Asp	TGG Trp 440	CTG Leu	GAC Asp	CCA Pro	GCC Ala	CGG Arg 445	ACC Thr	CGG Arg	1343
CTC Leu	AAC Asn	GGG Gly 450	GCC Ala	AAG Lys	ATG Met	AAG Lys	CAG Gln 455	CTC Leu	TTT Phe	AGC Ser	ATC Ile	CTG Leu 460	GAG Glu	GAG Glu	CTG Leu	1391
GCC Ala	ATG Met 465	GTG Val	ACC Thr	AGC Ser	CTG Leu	CGG Arg 470	CCA Pro	CCA Pro	GTA Val	CAG Gln	GCC Ala 475	AAC Asn	CCC Pro	GAC Asp	CTG Leu	1439

CTG	AGC	CTG	CTC	ACG	GTG	тст	CTG	GAT	CAG	TAT	CAG	ACG	GAG	GAT	GAG	1487
Leu 480	Ser	Leu	Leu	Thr	Val 485	Ser	Leu	Asp	Gln	Tyr 490	Gln	Thr	Glu	Asp	Glu 495	
CTG Leu	TAC Tyr	CAG Gln	CTG Leu	TCC Ser 500	CTG Leu	CAG Gln	CGG Arg	GAG Glu	CCG Pro 505	CGC Arg	TCC Ser	AAG Lys	TCC Ser	TCG Ser 510	CCA Pro	1535
ACC Thr	AGC Ser	CCC Pro	ACG Thr 515	AGT Ser	TGC Cys	ACC Thr	CCA Pro	CCA Pro 520	CCC Pro	CGG Arg	CCC Pro	CCG Pro	GTA Val 525	CTG Leu	GAG Glu	1583
GAG Glu	TGG Trp	ACC Thr 530	TCG Ser	GCT Ala	GCC Ala	AAA Lys	CCC Pro 535	AAG Lys	CTG Leu	GAT Asp	CAG Gln	GCC Ala 540	CTC Leu	GTG Val	GTG Val	1631
GAG Glu	CAC His 545	ATC Ile	GAG Glu	AAG Lys	ATG Met	GTG Val 550	GAG Glu	TCT Ser	GTG Val	TTC Phe	CGG Arg 555	AAC Asn	TTT Phe	GAC Asp	GTC Val	1679
GAT Asp 560	GGG Gly	GAT Asp	GGC Gly	CAC His	ATC Ile 565	TCA Ser	CAG Gln	GAA Glu	GAA Glu	TTC Phe 570	CAG Gln	ATC Ile	ATC Ile	CGT Arg	GGG Gly 575	1727
AAC Asn	TTC Phe	CCT Pro	TAC Tyr	CTC Leu 580	AGC Ser	GCC Ala	TTT Phe	GGG Gly	GAC Asp 585	CTC Leu	GAC Asp	CAG Gln	AAC Asn	CAG Gln 590	GAT Asp	1775
GGC Gly	TGC Cys	ATC Ile	AGC Ser 595	AGG Arg	GAG Glu	GAG Glu	ATG Met	GTT Val 600	TCC Ser	TAT Tyr	TTC Phe	CTG Leu	CGC Arg 605	TCC Ser	AGC Ser	1823
TCT Ser	GTG Val	TTG Leu 610	GGG Gly	GGG Gly	CGC Arg	ATG Met	GGC Gly 615	TTC Phe	GTA Val	CAC His	AAC Asn	TTC Phe 620	CAG Gln	GAG Glu	AGC Ser	1871
AAC Asn	TCC Ser 625	TTG Leu	CGC Arg	CCC Pro	GTC Val	GCC Ala 630	TGC Cys	CGC Arg	CAC His	TGC Cys	AAA Lys 635	GCC Ala	CTG Leu	ATC Ile	CTG Leu	1919
									CGA Arg							1967
CAC His	AAG Lys	CAG Gln	TGC Cys	AAG Lys 660	Asp	CGC Arg	CTG Leu	TCA Ser	GTT Val 665	GAG Glu	TGT Cys	CGG Arg	CGC Arg	AGG Arg 670	Ala	2015
				Leu					Pro						CAC His	2063
			His					Phe					Pro		AGG Arg	2111
		Ser					Ile					Val			GTG Val	2159
	Asp					Ile			TA	ATAG	ATGC	TG I	GGTT	GGAT	rc ·	2208
AAG	GACT	CAT	TCCI	GCCI	TG G	AGAA	AATA	C TI	CAAC	CAGA	GCA	GGGA	GCC	TGGG	GGTGTC	2268
GGG	GCAG	GAG	GCTG	GGGA	TG G	GGGT	'GGGA	TA T	GAGG	GTGG	CAT	GCAG	CTG	AGGG	CAGGGC	2328

CAGGGCTGGT	GTCCCTAAGG	TTGTACAGAC	TCTTGTGAAT	ATTTGTATTT	TCCAGATGGA	2388
ATAAAAAGGC	CCGTGTAATT	AACCTTC				2415
(2) INFORM	ATION FOR SI	EQ ID NO:5:				

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 728 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5: Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser His 10 Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser Pro Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly Arg Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val Gln 65 70 75 Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly Pro 105 Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val 185 Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala Glu 195 200 205

Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys Ala 210 215 220

Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp Ile

Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg Asn

Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His Leu 260 265 270

Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg Ser 280 Phe Cys Lys Ile Leu Phe Gln Asp Tyr His Ser Phe Val Thr His Gly 290 295 300 Cys Thr Val Asp Asn Pro Val Leu Glu Arg Phe Ile Ser Leu Phe Asn 310 Ser Val Ser Gln Trp Val Gln Leu Met Ile Leu Ser Lys Pro Thr Ala Pro Gln Arg Ala Leu Val Ile Thr His Phe Val His Val Ala Glu Lys 345 Leu Leu Gln Leu Gln Asn Phe Asn Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser Arg Leu Lys Glu Thr His Ser His Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr 385 390 395 400 Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys Val 405 410 415 Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala 425 Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu 490 Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr 505 Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu 520 Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly 585 Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn 610 615 620 Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu Gly

529

				•											
625				630					635					640	
Ile Tyr	Lys	Gln	Gly 645	Leu	Lys	Суѕ	Arg	Ala 650	Суѕ	Gly	Val	Asn	Суs 655	His	
Lys Glr	Cys	Lys 660	Asp	Arg	Leu	Ser	Val 665	Glu	Cys	Arg	Arg	Arg 670	Ala	Gln	
Ser Val	Ser 675	Leu	Glu	Gly	Ser	Ala 680	Pro	Ser	Pro	Ser	Pro 685	Met	His	Ser	
His His		Arg	Ala	Phe	Ser 695	Phe	Ser	Leu	Pro	Arg 700	Pro	Gly	Arg	Arg	
Gly Ser 705	Arg	Pro	Pro	Glu 710	Ile	Arg	Glu	Glu	Glu 715	Val	Gln	Thr	Val	Glu 720	
Asp Gly	Val	Phe	Asp 725	Ile	His	Leu									
(2) INF	יגאמט	rt Ori	₽ ∩₽	C E O	TD N	TO . 6 .									
(Z) INF	OMM.	LION	FOR.	SEQ	10 1	VO: 0:									
(i) SE(4	
		A) LE B) TY						S							
	((c) S1	RANI	EDNE	ESS:	sing								•	
	(1) TC	POLC	GY:	line	ear									
(ii) MOI	LECUL	E TY	PE:	DNA										
(ix) FE	ATURE	:												
		A) NA													
		3) LC	CATI	ON:	254.	208	33								
(vi) SE(MENC	יב חפ	ecp t	ייידר. סייידר	N	י ספי	ים איכ							
							_								
CGATTTC	ATT (CTCC	CTCC	C CA	CAGG	STCCC	TCI	CCCC	AAA	ATAT	TCCC	CAT (TTGI	CCTAG	60
CCCATCC	CCC I	AGACT	'ATCI	C AA	GGAC	CAGC	TGT	cccc	ACG	cccc	CGAC	CT C	CACI	AGGCC	120
TGTGCCA	ccc o	CTGC	CTGC	A GG	AAGA	CGCC	CGG	TCCC	GGG	CCGG	GTTA	GC C	CCAT	'GGGAA	180
CGGGGTT	CGG 1	rccga	'GCCC	G GI	GGGA	GGCI	ccc	GGAG	CGC	AGCC	TGGG	CC C	CAGCC	CACCC	240
cecece	GCG (CC A	TG G Met A	GCA G	GC A	ACC C	TG G Leu A	AC C	TG G	SAC A	AG C	ly (TGC A	CG hr	289
•							-					10		,	
GTG GAG Val Glu	GAG Glu 15	CTG Leu	CTC Leu	CGC Arg	GGG Gly	TGC Cys 20	ATC Ile	GAA Glu	GCC Ala	TTC Phe	GAT Asp 25	GAC Asp	TCC Ser	GGG Gly	337
AAG GTG	CGG	GAC	CCG	CAG	CTG	СТС	רפר	ልጥር	ሙጥር	כיייכ	אַתעכ	אשכי.	CAC	CCC	205
Lys Val	Arg	Asp	Pro	Gln	Leu	Val	Arg	Met	Phe	Leu	Met	Met	His	Pro	385
30					35					40					•
TGG TAC	ATC	CCC	TCC	TCT	CAG	CTG	GCG	GCC	AAG	CTG	ÇTC	CAC	ATC	TAC	433
Trp Tyr 45	TIE	Pro	ser	Ser 50	GIN	ren	Ата	AIA	Lys 55	Leu	Leu	Hıs	Ile	Tyr 60	
ראא ראא	TICC.	CCC.	220		220	mc-c	3.30	maa							
CAA CAA Gln Gln	Ser	Arg	AAG Lys	Asp	Asn	Ser	AAT Asn	Ser	Leu	Gln	GTG Val	AAA Lys	ACG Thr	TGC Cvs	481
		_	65	-				70				- -	75	- 4 -	

CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG GAG TTT GAC TTG

His	Leu	Val	Arg 80	Tyr	Trp	Ile	Ser	Ala 85	Phe	Pro	Ala	Glu	Phe 90	Asp	Leu	
									GAG Glu							577
									CTA Leu							625
									ACT Thr							673
_									TTT Phe 150						ATG Met	721
									GAG Glu							769
									GTG Val							817
									TCC Ser							865
									AAA Lys							913
									GTG Val 230							961
									GTG Val							1009
									CAC His							1057
									GAA Glu							1105
									GCA Ala							1153
									GAC Asp 310							1201
									CGG Arg							1249
									GAG Glu							1297

AGC Ser	CTG Leu 350	CGG Arg	CCA Pro	CCA Pro	GTA Val	CAG Gln 355	GCC Ala	AAC Asn	CCC	GAC Asp	CTG Leu 360	CTG Leu	AGC Ser	CTG Leu	CTC Leu	1345
ACG Thr 365	GTG Val	TCT Ser	CTG Leu	GAT Asp	CAG Gln 370	TAT Tyr	CAG Gln	ACG Thr	GAG Glu	GAT Asp 375	GAG Glu	CTG Leu	TAC Tyr	CAG Gln	CTG Leu 380	1393
TCC Ser	CTG Leu	CAG Gln	CGG Arg	GAG Glu 385	CCG Pro	CGC Arg	TCC Ser	AAG Lys	TCC Ser 390	TCG Ser	CCA Pro	ACC Thr	AGC Ser	CCC Pro 395	ACG Thr	1441
AGT Ser	TGC Cys	ACC Thr	CCA Pro 400	CCA Pro	CCC Pro	CGG Arg	CCC	CCG Pro 405	GTA Val	CTG Leu	GAG Glu	GAG Glu	TGG Trp 410	ACC Thr	TCG Ser	1489
GCT Ala	GCC Ala	AAA Lys 415	CCC Pro	AAG Lys	CTG Leu	GAT Asp	CAG Gln 420	GCC Ala	CTC Leu	GTG Val	GTG Val	GAG Glu 425	CAC His	ATC Ile	GAG Glu	1537
AAG Lys	ATG Met 430	GTG Val	GAG Glu	TCT Ser	GTG Val	TTC Phe 435	CGG Arg	AAC Asn	TTT Phe	GAC Asp	GTC Val 440	GAT Asp	GGG Gly	GAT Asp	GGC Gly	1585
CAC His 445	ATC Ile	TCA Ser	CAG Gln	GAA Glu	GAA Glu 450	TTC Phe	CAG Gln	ATC Ile	ATC Ile	CGT Arg 455	GGG Gly	AAC Asn	TTC Phe	CCT Pro	TAC Tyr 460	1633
	AGC Ser															1681
AGG Arg	GAG Glu	GAG Glu	ATG Met 480	GTT Val	TCC Ser	TAT Tyr	TTC Phe	CTG Leu 485	CGC Arg	TCC Ser	AGC Ser	TCT Ser	GTG Val 490	TTG Leu	GGG Gly	1729
GGG Gly	CGC Arg	ATG Met 495	GGC Gly	TTC Phe	GTA Val	CAC His	AAC Asn 500	TTC Phe	CAG Gln	GAG Glu	AGC Ser	AAC Asn 505	TCC Ser	TTG Leu	CGC Arg	1777
CCC Pro	GTC Val 510	GCC Ala	TGC Cys	CGC Arg	CAC His	TGC Cys 515	AAA Lys	GCC Ala	CTG Leu	ATC Ile	CTG Leu 520	GGC Gly	ATC	TAC Tyr	AAG Lys	1825
CAG Gln 525	GGC Gly	CTC Leu	AAA Lys	TGC Cys	CGA Arg 530	GCC Ala	TGT Cys	GGA Gly	GTG Val	AAC Asn 535	TGC Cys	CAC His	AAG Lys	CAG Gln	TGC Cys 540	1873
	GAT Asp															1921
CTG Leu	GAG Glu	GGG Gly	TCT Ser 560	GCA Ala	CCC Pro	TCA Ser	CCC Pro	TCA Ser 565	CCC Pro	ATG Met	CAC His	AGC Ser	CAC His 570	CAT His	CAC His	1969
CGC Arg	GCC Ala	TTC Phe 575	AGC Ser	TTC Phe	TCT Ser	CTG Leu	CCC Pro 580	CGC Arg	CCT Pro	GGC Gly	AGG Arg	CGA Arg 585	GGC Gly	TCC Ser	AGG Arg	2017
CCT Pro	CCA Pro 590	GAG Glu	ATC Ile	CGT Arg	GAG Glu	GAG Glu 595	GAG Glu	GTA Val	CAG Gln	ACG Thr	GTG Val 600	GAG Glu	GAT Asp	GCG	GTG Val	2065
TTT Phe 605	GAC Asp	ATC Ile	CAC His	TTG Leu	TAA1	TAGAT	GC 1	GTGC	STTGO	GA TO	CAAGO	SACTO	ATT	CCTC	SCCT	2120

TTAACCTTC					,	2309
GGTTGTACAG	ACTCTTGTGA	ATATTTGTAT	TTTCCAGATG	GAATAAAAAG	GCCCGTGTAA	2300
TGGGGGTGGG	ATATGAGGGT	GGCATGCAGC	TGAGGGCAGG	GCCAGGGCTG	GTGTCCCTAA	2240
TGGAGAAAAT	ACTTCAACCA	GAGCAGGGAG	CCTGGGGGTG	TCGGGGCAGG	AGGCTGGGGA	2180

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 609 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu 15

Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro Trp Tyr Ile Pro 45

Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr Gln Gln Ser Arg 65

Asp Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys His Leu Val Arg 80

Tyr Trp Ile Ser Ala Phe Pro Ala Glu Phe Asp Leu Asn Pro Glu Glu Leu 95

Ala Glu Gln Ile Lys Glu Leu Lys Ala Leu Leu Asp Gln Gln Gly Asn 100

Arg Arg His Ser Ser Leu Ile Asp Ile Asp Ser Val Pro Trp Tyr Lys 112

Trp Lys Arg Gln Val Thr Gln Arg Asn Pro Val Gly Gln Lys Lys Arg 130

Lys Met Ser Leu Leu Phe Asp His Leu Glu Pro Met Glu Leu Ala Glu His Leu Thr Tyr Leu Glu Tyr Arg Ser Phe Cys Lys Ile Leu Phe Gln 175

Asp Tyr His Ser Phe Val Thr His Gly Cys Thr Val Asp Asn Pro Val 180

Leu Glu Arg Phe Ile Ser Leu Phe Asn Ser Val Ser Gln Trp Val Gln 195

Leu Met Ile Leu Ser Lys Pro Thr Ala Pro Gln Arg Ala Leu Val Ile

Thr His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe

Asn Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser

Leu

Arg Leu Lys Glu Thr His Ser His Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr Ala Thr Gly Asn Tyr Gly Asn 280 Tyr Arg Arg Arg Leu Ala Ala Cys Val Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala Leu Gln Leu Ala Leu Pro Asp Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala Met Val Thr Ser Leu Arg Pro 345 Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu Lys Met Val Glu 425 Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met 470 Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys 505 Arg His Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu Asp Gly Val Phe Asp Ile His - 72 -

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 832 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 11..733

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

(3117) 0120		002 20 110		
Met			GC CTG TGC CGG C rg Leu Cys Arg L 10	
		Leu Leu Gly	GCG GCC GCC GGG (Ala Ala Ala Gly (25	
TCC AGA CCC AG Ser Arg Pro So 30	GT ACT TAT TAT er Thr Tyr Tyr 35	GAA CTG TTG Glu Leu Leu	GGG GTG CAT CCT (Gly Val His Pro (40	GGT GCC 145 Gly Ala 45
			TCC AAG TCC AAA Ser Lys Ser Lys	
His Pro Asp A			CTG CAC AGC CGC Leu His Ser Arg 75	
			CGT GAG CAG AGC Arg Glu Gln Ser 90	
			CCC CCA AAG TCT Pro Pro Lys Ser 105	
			ACA CAC AGC TCC Thr His Ser Ser 120	
			CAC AGC GTG AGG His Ser Val Arg	
Gly Pro Gln L			CAA AAC AAA CAA Gln Asn Lys Gln 155	
			ATG GGC CTG CAC Met Gly Leu His 170	
		Met His Leu	AAC TTC ATG GAT Asn Phe Met Asp 185	

GAT Asp 190	CGG Arg	ATC Ile	ATC Ile	ACA Thr	GCC Ala 195	TTC Phe	TAC Tyr	AAC Asn	GAA Glu	GCC Ala 200	CGG Arg	GCA Ala	CGG Arg	GCC Ala	AGG Arg 205		625
GCC Ala	AAC Asn	AGA Arg	GGC Gly	ATC Ile 210	CTT Leu	CAG Gln	CAG Gln	GAG Glu	CGA Arg 215	CAA Gln	CGG Arg	CTA Leu	GGG Gly	CAG Gln 220	CGG Arg		673
CAG Gln	CCG Pro	CCA Pro	CCA Pro 225	TCC Ser	GAG Glu	CCA Pro	ACC Thr	CAA Gln 230	GGC Gly	CCC Pro	GAG Glu	ATC Ile	GTG Val 235	CCC Pro	CGG Arg		721
	GCC Ala			TGA	GGG	GCTC	ACC!	rgga:	rgg (GCC'	rgca	GT G	CGTT(CCCG	2		773
TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC GCAATAAAGT GATTCGCAG 83										832							
(2)	INFO	ORMA	rion	FOR	SEQ	ID i	10:9	•									
٠	ı	(i) :	(B)	LEI TYI	NGTH:	: 241		ino a id	: acids	5							
	(i	ii) 1	MOLEC	CULE	TYPE	iq :E	rotei	in									
	()	ci) s	SEQUE	ENCE	DESC	CRIPT	NOI1	SE	Q ID	NO:):						
Met 1	Pro	Pro	Leu	Leu 5	Pro	Leu	Arg	Leu	Cys 10	Arg	Leu	Trp	Pro	Arg 15	Asn		
Pro	Pro	Ser	Arg 20	Leu	Leu	Gly	Ala	Ala 25	Ala	Gly	Gln	Arg	Ser 30	Arg	Pro		
Ser	Thr	Tyr 35	Tyr	Glu	Leu	Leu	Gly 40	Val	His	Pro	Gly	Ala 45	Ser	Thr	Glu		
Glu	Val 50	Lys	Arg	Ala	Phe	Phe 55	Ser	Lys	Ser	Lys	Glu 60	Leu	His	Pro	Asp		
Arg 65	Asp	Pro	Gly	Asn	Pro 70	Ser	Leu	His	Ser	Arg 75	Phe	Val	Glu	Leu	Ser 80		
Glu	Ala	Tyr	Arg	Val 85	Leu	Ser	Arg	Glu	Gln 90	Ser	Arg	Arg	Ser	Tyr 95	Àsp		
Asp	Gln	Leu	Arg 100	Ser	Gly	Ser	Pro	Pro 105	Lys	Ser	Pro	Arg	Thr 110	Thr	Val		
His	Asp	Lys 115	Ser	Ala	His	Gln	Thr 120	His	Ser	Ser	Trp	Thr 125	Pro	Pro	Asn		
Ala	Gln 130	Tyr	Trp	Ser	Gln	Phe 135	His	Ser	Val	Arg	Pro 140	Gln	Gly	Pro	Gln		
Leu 145	Arg	Gln	Gln	Gln	His 150	Lys	Gln	Asn	Lys	Gln 155	Val	Leu	Gly	Tyr	Cys 160		
Leu	Leu	Leu	Met	Leu 165	Ala	Gly	Met	Gly	Leu 170	His	Tyr	Ile	Ala	Phe 175	Arg		
Lys	Val	Lys	Gln 180	Met	His	Leu	Asn	Phe	Met	Asp	Glu	Lys	Asp	Arg	Ile		

Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg Ala Asn Arg

WO 98/53061 PCT/AU98/00380

- 74 -

200 205 195 Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg Gln Pro Pro 215 Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg Gly Ala Gly 235 230 Pro SEQ ID Nos: 10-18 25-36 (2) INFORMATION FOR SEQ ID NO:7: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 300 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 170..300 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7: CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG 60 CCCATCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCGACCT CCACTAGGCC 120 TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAG CCC CAT 175 Pro His GGG AAC GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC 223 Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser 10 CTG GGC CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC 271 Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp 300 CTG GAC AAG GGC TGC ACG GTG GAG GAG CT Leu Asp Lys Gly Cys Thr Val Glu Glu Leu (2) INFORMATION FOR SEQ ID NO:8: (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids(B) TYPE: amino acid(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

PCT/AU98/00380

- 75 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Pro His Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu

1 10 15

Arg Ser Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr 20 25 30

Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu 35

- (2) INFORMATION FOR SEQ ID NO:9:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGATCCCCC TGGTC

15

- (2) INFORMATION FOR SEQ ID NO:10:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Asp Val Asp Glu Glu Asp Glu Val Glu Asp Ile Glu Phe 1 5 10

- (2) INFORMATION FOR SEQ ID NO:11:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Asp Val Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe 1 5 10

- (2) INFORMATION FOR SEQ ID NO:12:
 - (i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 13 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asp His Asp Arg Asp Gly Phe Ile Ser Gln Glu Glu Phe

- (2) INFORMATION FOR SEQ ID NO:13:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met

- (2) INFORMATION FOR SEQ ID NO:14:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asp Val Asp Met Asp Gly Gln Ile Ser Lys Asp Glu Leu

- (2) INFORMATION FOR SEQ ID NO:15:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids

 - (B) TYPE: amino acid(C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

His Phe Val His Val Ala Glu Lys Leu Gln Leu Gln Asn Phe Asn

Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser Arg

- 77 -

Leu Lys Glu Thr His 35

- (2) INFORMATION FOR SEQ ID NO:16:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Lys Phe Val His Val Ala Lys His Leu Arg Lys Ile Asn Asn Phe Asn 1 5 10 15

Thr Leu Met Ser Val Val Gly Gly Ile Thr His Ser Ser Val Ala Arg 20 25 30

Leu Ala Lys Thr Tyr

- (2) INFORMATION FOR SEQ ID NO:17:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys Arg His 1 5 10 15

Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg 20 25 30

Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu Ser Val 35 40 45

Glu Cys 50

- (2) INFORMATION FOR SEQ ID NO:18:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 amino acids
 - (B) TYPE: amino acid(C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
 - His Asn Phe His Glu Thr Thr Phe Leu Thr Pro Thr Thr Cys Asn His

	1		•		5					10					15		
	Cys	Asn	Lys	Leu 20	Leu	Trp	Gly	Ile	Leu 25	Arg	Gln	Ġly	Phe	Lys 30	Cys	Lys	
	Asp	Cys	Gly 35	Leu	Ala	Val	His	Ser 40	Cys	Cys	Lys	Ser	Asn 45	Ala	Val	Ala	
	Glu	Cys 50															
(2) INFORMATION FOR SEQ ID NO:19:																	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 15 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 																	
	(ii)	MOL	ECUL	Е ТҮ	PE:	DNA											
	(xi)	SEQ	UENC	Ė DE	SCRI	PTIO	N: S	EQ I	D NO	:19:							
GGG	TCCC	CC T	GGTC														15
(2)	INFO	RMAT	ION	FOR	SEQ	ID N	0:20	:									
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear																
	(ii)	MOL	ECUL	Е ТҮ	PE:	DNA											
	(xi)	SEQ	UENC	E DE	SCRI	PTIC	N: S	EQ I	D NO	:20:							
GAA'	rtcgg	CA C	GAGC	CGAC	G G												21
(2)	INFO	RMAT	NOI	FOR	SEQ	ID N	10:21	. :									
	(i)	(A (E (C	l) LE 3) TY :) ST	NGTH PE:	i: 78 nuc] EDNE	bas eic SS:	STIC se pa acid sing ar	irs l									
	(ii)	MOI	ECUI	LE TY	PE:	DNA											
	(xi)	SEÇ	QUENC	CE DE	ESCR	PTIC	on: s	SEQ I	D NO):21:	:		•				
ATG	GAGC	GA A	GCT	BATCT	rc co	GAGG	AGGA	CTC	CCCC	GGG	CAGO	TGGA	TC C	CGCAC	GCCCA	VC	60
ccc	GCGC	GG C	CGGC	CATG													78
(2)	INFO	RMAT	пои	FOR	SEQ	ID I	NO : 22	2:									

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Pro Gly Ala Ala Gly

Ser Ala Ala His Pro Ala Pro Ala Ala Met

- (2) INFORMATION FOR SEQ ID NO:23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGATCCGCAG CCCACCCCGC GCCGGCGGCC ATG

33

- (2) INFORMATION FOR SEQ ID NO:24:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met

- (2) INFORMATION FOR SEQ ID NO:25:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA

20

GTCATAGTCT GTCTCCTACT

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
GGAC.	AAAGTG TGTGATGAAC C	21
(2)	INFORMATION FOR SEQ ID NO:26:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
CTCA	TCCTCC GTCTGATACT G	21
(2)	INFORMATION FOR SEQ ID NO:27:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
GTAG	SATGTGG ATCAGCTTGG	20
(2)	INFORMATION FOR SEQ ID NO:28:	•
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
AGGT	rggagaa tggtcaagg	19
(2)	INFORMATION FOR SEQ ID NO:29:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	

WO 98/53061 PCT/AU98/00380

- 81 -

(2)	INFORMATION FOR SEQ ID NO:30:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:	
ACAT	AGACAG CGTGCCTACC	20
(2)	INFORMATION FOR SEQ ID NO:31:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:	:
TACA	ACCTTA GGGACACCAG	20
(2)	INFORMATION FOR SEQ ID NO:32:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
TGCI	CGAGCCT GCTCACGGTG	20
(2)	INFORMATION FOR SEQ ID NO:33:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
CAAC	STGAACA GCACGTCC	18
(2)	INFORMATION FOR SEQ ID NO:34:	
	(i) SEQUENCE CHARACTERISTICS:	

WO 98/53061 PCT/AU98/00380

- 82 -

	(A) LENGTH: 21 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
GAC:	PATCTCA AGGACCAGCT G	21
(2)	INFORMATION FOR SEQ ID NO:35:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
GGT:	PCGGTCC GAGCCCGG	18
(2)	INFORMATION FOR SEO ID NO:36:	
(2)	-	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:	
GGA	GCGATAC TCCAAGTAGG T	21
(2)	INFORMATION FOR SEQ ID NO:37:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
AGC	GGGCCAG GCCCCTTC	18
	TWO THE TOTAL TOTAL TO ME TO M	
(2)	INFORMATION FOR SEQ ID NO:38:	

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
CATCCTGGTC CAATGCGCTC	20
(2) INFORMATION FOR SEQ ID NO:39:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:	
GCACTGAGGA AGTTAAACGA GC	22
(2) INFORMATION FOR SEQ ID NO:40:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:	
GCTCGTTTAA CTTCCTCAGT GC	22
(2) INFORMATION FOR SEQ ID NO:41:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: DNA	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:	
GCTCAGCTCC ACAAAGCGGC T	. 21
(2) INFORMATION FOR SEQ ID NO:42:	

	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:42:	
ACCA	GCTC	CG CTCAGGTAG	19
(2)	INFO	RMATION FOR SEQ ID NO:43:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:43:	
TCCA	GGAG	CT GTGTGTTTGG	20
(2)	INFO	RMATION FOR SEQ ID NO:44:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:44:	
CCAC	TTTC	AC AGCGTGAGG	19
(2)	INFO	RMATION FOR SEQ ID NO:45:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: DNA	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:45:	

19

CLAIMS:

- 1. An isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
- 2. An isolated nucleic acid molecule according to claim 1 wherein the regulator comprises a zinc finger domain of an (HC₃)₂ type.
- 3. An isolated nucleic acid molecule according to claim 2 wherein the sequence of nucleotides or complementary sequence of nucleotides is selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
- 4. An isolated nucleic acid molecule according to claim 1 wherein said gene regulator is a guanine nucleotide exchange factor (GEF) or a derivative thereof.
- 5. An isolated nucleic acid molecule according to claim 4 wherein the sequence of nucleotides is selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the

nucleotide sequence set forth in (i), (ii) or (iii).

- 6. An isolated nucleic acid molecule according to claim 1, wherein said gene regulator is a heat shock protein or is a heat shock binding protein or a derivative thereof.
- 7. An isolated nucleic acid molecule according to claim 6, wherein the sequence of nucleotides is selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
- 8. A genetic construct comprising a vector portion and a gene portion comprising a regulator of gene expression or a derivative thereof.
- 9. A genetic construct according to claim 8 wherein the gene portion comprises a zinc finger domain of (HC₃)₂ type.
- 10. A genetic construct according to claim 9 wherein the gene portion comprises a nucleotide sequence selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

WO 98/53061 PCT/AU98/00380

- 11. A genetic construct according to claim 8 wherein said gene portion is a nucleotide exchange factor (GEF) or derivative thereof.
- 12. A genetic construct according to claim 11 wherein the gene portion comprises a nucleotide sequence selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
- 13. A genetic construct according to claim 8 wherein the gene portion is a heat shock protein or a derivative thereof or a heat shock binding protein or derivative thereof.
- 14. A genetic construct according to claim 13 wherein the gene portion comprises a nucleotide sequence selected from:
- (i) a nucleotide sequence set forth in SEO ID NO:8;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
- 15. A nucleic acid molecule encoding a gene regulator having the identifying characteristics of a molecule selected from MCG4, MCG7 and MCG18 having respective amino acid sequences of SEQ ID NO:3, SEQ ID NO:5 or 7 and SEQ ID NO:9.

- 16. A method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg4 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.
- 17. A method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.
- 18. A method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.
- 19. A method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg7 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.
- 20. A method of detecting a condition caused or facilitated by an aberration in *mcg7*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.
- 21. A method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

- 22. A method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg18 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.
- 23. A method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.
- 24. A method for detecting MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

FIGURE 1

TCAGTAAACA CAGAGACTGG GGATCGATC	ATG GGG CTT ? Met Gly Leu (TGT AAG TGC CCC Cys Lys Cys Pro 5	AAG 53 Lys
AGA AAG GTG ACC AAC CTG TTC TGC Arg Lys Val Thr Asn Leu Phe Cys 10 15	TTC GAA CAT (CGG GTC AAC GTC Arg Val Asn Val 20	TGC 101 Cys
GAG CAC TGC CTG GTA GCC AAT CAC Glu His Cys Leu Val Ala Asn His 25 30	GCC AAG TGC AAA Lys Cys 35	ATC GTC CAG TCC Ile Val Gln Ser	TAC 149 Tyr 40
CTG CAA TGG CTC CAA GAT AGC GAC Leu Gln Trp Leu Gln Asp Ser Asp 45			
AAC ATA CCC CTG GCC AGC CGA GAG Asn Ile Pro Leu Ala Ser Arg Glu 60	Thr Thr Arg	Leu Val Cys Tyr 70	Asp
CTC TTT CAC TGG GCC TGC CTC AAT Leu Phe His Trp Ala Cys Leu Asn 75 80	Glu Arg Ala .	Ala Gin Leu Pro 85	Arg
AAC ACG GCA CCT GCC GGC TAT CAG Asn Thr Ala Pro Ala Gly Tyr Gln 90 95	. Cys Pro Ser	Cys Asn Gly Pro 100	Ile
TTC CCC CCA ACC AAC CTG GCT GGC Phe Pro Pro Thr Asn Leu Ala Gly 105 110	Pro Val Ala 115	Ser Ala Leu Arg	Glu 120
AAG CTG GCC ACA GTC AAC TGG GCC Lys Leu Ala Thr Val Asn Trp Ala 125	Arg Ala Gly 130	Leu Gly Leu Pro 135	Leu
ATC GAT GAG GTG GTG AGC CCA GAG Ile Asp Glu Val Val Ser Pro Glu 140	Pro Glu Pro 145	Leu Asn Thr Ser 150	Asp
TTC TCT GAC TGG TCT AGT TTT AAT Phe Ser Asp Trp Ser Ser Phe Asr 155	n Ala Ser Ser	Thr Pro Gly Pro 165	Glu
GAG GTA GAC AGC GCC TCT GCT GCC Glu Val Asp Ser Ala Ser Ala Ala 170 175	C CCA GCC TTC a Pro Ala Phe	TAC AGC CGA GCC Tyr Ser Arg Ala 180	CCC 581 Pro
CGG CCC CCA GCT TCC CCA GGC CGC Arg Pro Pro Ala Ser Pro Gly Arg 185	g Pro Glu Gln 195	. His Thr Val Ile	e His 200
ATG GGC AAT CCT GAG CCC TTG AC Met Gly Asn Pro Glu Pro Leu Th 205	T CAC GCC CCT r His Ala Pro 210	AGG AAG GTG TAT Arg Lys Val Tyr 219	: Asp

Thr Arg Asp Asp Asp Asp Thr Pro Gly Leu His Gly Asp Cys Asp Asp 220 225 233 GAC AAG TAC CGA CGT CGG CCG GCC TTG GGT TGG CTG GCC CGG CTG CT																	
GAC AAG TAC CGA CGT CGG CCG GCC TTG GGT TGG CTG GCC CGG CTA ASP Lys Tyr Arg Arg Arg Pro Ala Leu Gly Trp Leu Ala Arg Leu Leu 235 AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 300 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCCCA AGCCCTCGAC CCCTCGCCCC 1202 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1203	ACG	CGG	GAT	GAT	عدد	CGG	ACA	CCA	SGC	CTC	CAT	GGA	GAC	₹.	BAC	SAT .	. ١٠٠٠ ناري .
GAC AAG TAC CGA CGT CGG CCG GCC TTG GGT TGG CTG GCC CGG CTA ASP Lys Tyr Arg Arg Arg Pro Ala Leu Gly Trp Leu Ala Arg Leu Leu 235 240 245 AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTC CAG CGG Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 255 260 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 285 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCCCA AGCCCTCGAC CCCTCGCCCC 1206 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1206 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1206 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1206	Thr	Arg	Asp	àsp	Asp	Arg	Thr	Pro	Gly	Leu	Hıs	Gly	Asp	Сув	Asp	Asp	
ASP LYS TYR ARG ARG ARG PRO Ala Leu Gly Trp Leu Ala Arg Leu Leu 235 240 245 AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG 821 ARG Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 260 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT 869 Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1200				220					225					230			
ASP LYS TYR ARG ARG ARG PRO Ala Leu Gly Trp Leu Ala Arg Leu Leu 235 240 245 AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG 821 ARG Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 260 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT 869 Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1200																	
AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 290 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 300 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGACTAAGA CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCCCCC TCAAGCCTAA GACACTAAGA GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCGAC CCCTCGCCCC 1202																	773
AGG AGC CGG GCT GGG TCT CGG AAG CGG CCG CTG ACC CTG CTC CAG CGG Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 300 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	Asp	Lys		Arg	Arg	Arg	Pro		Leu	Gly	Trp	Leu			Leu	Leu	•
Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 300 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202			235			•		240					245				
Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg 250 GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 300 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	100		-	~~~	~~~	mom	-			ccc	CTC	N.C.C	CTC	CT-	030	ccc	001-
GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202																	021
GCG GGG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG GCC CTC CTT Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu 265 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	Arg		Arg	ALG	GIA	Ser	_	Lys	Arg	PIO	Leu		Leu	Leu	3111	ALG	
Ala Gly Leu Leu Leu Leu Cly Leu Leu Gly Phe Leu Ala Leu Leu 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC 917 Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCGAC CCCTCGCCCC 1202		250					200					200					
Ala Gly Leu Leu Leu Leu Cly Leu Leu Gly Phe Leu Ala Leu Leu 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC 917 Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCGAC CCCTCGCCCC 1202	GCG	GGG	CTG	СТС	СТА	CTC	TTG	GGA	CTG	CTG	GGC	TTC	CTG	GCC	CTC	CTT	869
265 270 275 280 GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC AGC GAT CCC AAC 917 Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202																	
Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202		,						,									
Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn 285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA 962 Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	•				•												
285 290 295 CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	GCC	CTC	ATG	TCT	CGC	CTA	GGC	CGG	GCC	GCA	GCT	GAC	AGC	GAT	CCC	AAC	917
CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG GGC CCC TCC TGA Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	Ala	Leu	Met	Ser	Arg	Leu	Gly	Arg	Ala	Ala	Ala	Asp	Ser	Asp	Pro	Asn	
Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202					285			,		290					295		
Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser 300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202																	•
300 305 310 GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202																	962
GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT CTGTGGAGGA GAGGCGGGGT 1022 AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA 1082 CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	Leu	Asp	Pro		Met	Asn	Pro	His		Arg	Val	Gly	Pro		•		•
AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202				300					305.					310			٠.
AATGGGGAGG CTGAGGGCAC CTCTTCACTG CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	ccc	~~~~	יייי י	nancar.				~~~ × ′		rcrc	ىلىلتات	CTG	rcca	302	azaa	CCCCT	1022
CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	GCC		iGC .	11610	36617	46 6	LAG	CIAC	J GA.	1919	3011	CIG.	COA	307	0.000	.00001	
CCCCAGACCC AAAGCCAAGT CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG 1142 GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202	AAT	GGGG#	AGG (CTGAC	3GGC/	AC C	CTTC	CACTO	CCC	CTC'	rccc	TCA	AGCC'	TAA	GACA	CTAAGA	1082
GGTCAAGCAT TTGTCTTGAC TTGCTTTCTC CCGGGTCTCC AGCCTCCGAC CCCTCGCCCC 1202																	
	CCC	CAGAC	CCC A	AAAG	CAA	GT. C	CACC	AGAG?	r GG	CTCG	CAGG	CCA	GCC'	TGG	AGTC	CCCGTG	1142
1242	GGT	CAAGO	CAT	TTGT	CTTG	AC T	rgct:	TTCT	2 000	GGT	CTCC	AGC	CTCC	GAC	CCCT	CGCCCC	1202
1949																	,
ATGAAGGAGC TGGCAGGTGG AAATAAACAA CAACTTTATT 1242	ATG	AAGG2	AGC	rggc <i>i</i>	AGGT	GG A	LATA	AACA	A CA	ACTT'	TATT						1242

3/32

Figure 2

gb|AA155210|AA155210 mr98e01.rl Stratagene mouse embryonic carcinoma (#937317) Mus musculus cDNA clone 605496 5

2 MGLCKCPKRKVTNLFCFEHRVNVCEHCLVANHAKCIYQSYLQWLQDSDYNPNCRLCNIPL 60 Query:

MGLCKCPKRKVTNLFCFEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCN PL 38 MGLCKCPKRKVTNLFCFEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCNTPL 277 Sbjct:

Figure 3

dbj|D75913|CELK111G3F C.elegans cDNA clone ykl11g3 : 5' end, single read.

7 PKRKVINLFCFEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCNIPLASRETT 66 PKRKVTNLF +EHRVNVCE LV NH C+VQSYL WL D DY+PNC LC L +T Query:

1 PKRKVINLFXYEHRVNVCELXLVDNHPNCVVQSYLTWLTDQDYDPNCSLCXTTLXEGDTI 180 Sbjct:

67 RLVCYDLFHWACLNERAAQLPRNTAPAGYQCP 98 98 PSCNGPIFPPNQ 109 Query:

RL C L HW C +E P TAP GY+CP P C+ +FPP+Q

181 RLNCLHILLHWKCFDEWXGNFPDTTAPXGYRCP 276 275 PCCSQEVFPPDQ 310 Sbjct:

Figure 4

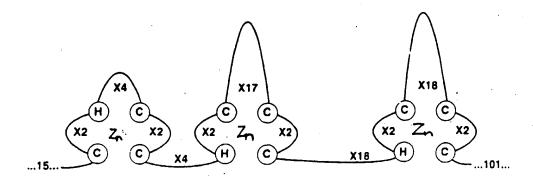


Figure 5

sp|P46580|YLB5_CAEEL HYPOTHETICAL 146.8 KD PROTEIN C34E10.5 IN CHROMOSOME III gi|500728 (U10402) C34E10.5 gene product [Caenorhabditis elegans]

Query: 56 CNIPLASRETTRLVCYDLFHWACLNERAAQLPRNTAPAGYQCPSC 100 C+I L ++ + L C LF W C+ E A ++ + + CP C Sbjct: 1222 CSICLENKNPSALFCGHLFCWTCIQEHAVAATSSASTSSARCPQC 1266

Figure 6

gi|703468 (L29051) homologous to GATA-binding transcription factor [Schizosaccharomyces pombe]

Query: 35 CIVQSYLQWLQDSDYNPNCRLCNI 58 C + +W +D NP C C + Sbjct: 175 CATTNTPKWRRDESGNPICNACGL 198

Query: 162 SSTPGPEEVDSASAAPAFYSQAPRPPASPGRPEQHTVIHMCNPEPLTHAPRKVYDTRDDD 221
+S PEE S S S P+ SP + +Q +I P +V + D
Sbjct: 441 ASLLAPEEPPSNSDKQPSMSNGPKSEVSPSQSQQAPLIQSSTSPVSLQFPPEVQGSNVDK 500

Query: 222 RTPGLH 227 R L+ Sbjet: 501 RNYALN 506

Figure 7



6/32

gb|AA074703|AA074703 zm76g07.rl Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone 531612 5' Length = 417

Plus Strand HSPs:

```
Score = 818 (226.0 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103
      Identities = 206/259 (79%), Positives = 206/259 (79%), Strand = Plus / Plus
                     446 GGCCTCCCTCTGATCGATGAGGTGGTGAGCCCCAGAGCCCCTCAACACGTCTGAC 505
   Query:
                            <u> 11 minimumini i madiminimum i mar</u>
   Sbjct:
                       49 GGGCTCCCTCTGATCGATGAGGTGATAAGCCCAGAGCCCGAGCCCCTCAATTCCTCAGAC 108
                    506 TYCTCTGACTGGTCTAGTTTTAATGCCAGCAGTACCCCTGGACCAGAGGAGGTAGACAGC 565
   Query:
                    Sbjct:
                    566 GCCTCTGCTGCCCCAGCCTTCTACAGCCAGCCCCCGGCCCCCAGCTTCCCCAGGCCGG 625
   Query:
                                  A THE REPORT OF THE PARTY OF TH
                    169 ACTCCATCTGCACCTGCTTTCTATAGCCAGGCTCCCCGCCCTCCTCCCCCAAGCCGT 228
  Sbjct:
                   626 CCCGAGCAGCACACAGTGATCCACATGGGCAATCCTGAGCCCTTGACTCACGCCCCTAGG 685
  Query:
                           229 CCCGAGCACACACACTCATACACATGGGGAGTACTGAAGCCCTGGCACACGCCCCAAGG 288
  Sbjct:
  Query:
                   686 AAGGIGTATGATACGCGGG 704
                           11 11 11 11 11 11 11
                   289 AAAGTATATGACACCCGG 307
   Score = 230 (63.6 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103
   Identities = 50/55 (90%), Positives = 50/55 (90%), Strand = Plus / Plus
 Ouerv:
                  398 GCACTGAGAGAGAGCTGGCCACAGTCAACTGGGCCCGGGCAGGACTGGGCCTCC 452
                          Sbict:
                      2 GCACTGAGAGAAAAGCTAGCCACAGTCAACTTGGCCCGGGCAGGACTGGGCTCCC 56
  Score = 175 (48.4 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103
  Identities = 39/44 (88%), Positives = 39/44 (88%), Strand = Plus / Plus
Query:
                  767 GCCTTGGGTTGGCTGGCCGGGCTGCTAAGGAGCCGGGCTGGGTC 810
                 373 GCTCTGGGCTGGCCCCAGCTGCTCAGGAGCCGGGCTGGGTC 416
Sbjct:
 Score = 139 (38.4 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103
Identities = 31/35 (88%), Positives = 31/35 (88%), Strand = Plus / Plus
Ouerv:
                 731 GGAGACTGTGACGATGACAAGTACCGACGTCGGCC 765
                         336 GGAGACTGTGATGATGACAAATACCGCCGCCGGCC 370
Sbict:
 Score = 133 (36.8 bits), Expect = 6.1e-103, Sum P(5) = 6.1e-103
 Identities = 29/32 (90%); Positives = 29/32 (90%), Strand = Plus / Plus
Query:
                 701 CGGGATGATGACCGGACACCAGGCCTCCATGG 732
                         1414144114444444 1114 1 1144
                 305 COGGATGATGACCGGACAGCAGGCATTCATGG 336
```

Figure 8 continued

```
gb|AA134788|AA134788 zm81g02.rl Stratagene neuroepithelium (#937231)
        Homo sapiens cDNA clone 532082 5'
        Length = 368
 Plus Strand HSPs:
Score = 563 (155.6 bits), Expect = 3.8e-87, Sum P(3) = 3.8e-87
Identities = 147/190 (77%), Positives = 147/190 (77%), Strand = Plus / Plus
      498 CGTCTGACTTCTCTGACTGGTCTAGTTTTAATGCCAGCAGTACCCCTGGACCAGAGGAGG 557
Ouerv:
         Sbict:
      558 TAGACAGCGCCTCTGCTGCCCCAGCCTTCTACAGCCAGGCCCCCGGCCCCCAGCTTCCC 617
Ouerv:
         Sbjct:
      618 CAGGCCGGCCCGAGCAGCACACAGTGATCCACATGGGCAATCCTGAGCCCTTGACTCACG 677
Query:
         223 CAAGCCGTCCCGAGCACACACACTCATACACATGGGGAGTACTGAAGCCCTGGCACACG 282
Sbjct:
      678 CCCCTAGGAA 687
Query:
         1111 11111
      283 CCCCAAGGAA 292
Sbjct:
Score = 454 (125.4 bits), Expect = 3.8e-87, Sum P(3) = 3.8e-87
Identities = 94/98 (95%), Positives = 94/98 (95%), Strand = Plus / Plus
      Query:
       Sbjct:
      458 ATCGATGAGGTGGTGAGCCCAGAGCCCCGAGCCCCTCAA 495
Query:
      Score = 219 (60.5 bits), Expect = 3.8e-87, Sum P(3) = 3.8e-87
Identities = 51/60 (85%), Positives = 51/60 (85%), Strand = Plus / Plus
      702 GGGATGATGACCGGACACCAGGCCTCCATGGAGACTGTGACGATGACAAGTACCGACGTC 761
Query:
         309 GGATTGATGACCGGACAGCAGGCATTCATGGAGACTGTGATGATGACAAATACCGCCGCC 368
Sbjct:
```

Figure 9

FIGURE 10

MCG4 MCG4 3. [229]	MGLCKCPKR ASRETTRLV	K VTNLFCFEHI C YDLFHWACLI	r vnvcehclv V eraaqlprn	A NHAKCIVQS T APAGYQCPS	Y LOWLODSDYI C NGPIFPPTNI	PNCRLCNIPL 60 AGPVASALRE 120
5.						***x>
[74]					·	****>
	130	140	150) 160	170	180
MCG4 1.	KLATVIMARI	GLGLPLIDEV				DSASAAPAFY
[372]		20 	30 i******	40	50 *tt*sva***	60 a*tps****>
2. [243]		· · · · ·	30	40	50	60
(243]			aqs*s*sip		*tt*svq**r	a*tps****>
3. [229]	10	20	30 i******s		50 chhhlcarge	60 sqh*icac*l>
		s		-		s
5.	10	-	30	40	50	60
[74]	******X***	****smr**a	q**s*-sipq	tslig-pal-	mppp*lckrr	ep*lhlxlli>
	R 190	•	•	•	230	240
MCG4	SPAPRPPASP	GRPEQHTVIH	MCNPEPLTHA	PRKVYDTRDD	DRTPGLHGDC	DDDKYRRRPA
1.				!		
[372]	. 70	80 . s*******	90 **st*a*a**	100	110 *srhswetvm	120
2. [243]	V /0	80	90			
(243)	9	S*******	sc-a-a	i		
3. [229]	70	80	90		110	120
4.	gsp*sslpk* 70	80	90	100	110	120
[86]	p*sslpk*	s*a-a*sht*	gey*s*g*rp	kesi*h*gmm h	tgqqafm***	*********C>
5.	70	80	90	100	110	•
[74]	arl*allppq	av*sstqsyt	w*vlk*w-*t	å ådåk*m****	***a*i**>	
6.				100		
[38]			•t	*q*******>		
	250	260	270	280	290	300
MCG4 1.	LOWLARLLES	RAGSRKRPLT	LLQRAGLLLL	LGLLGFLALL	almsrigraa .	ADSDPNILDPL
[372]	130 *****q****	****>	ı			
4. [86]	S*-**>		•			
	310					
MCG4	MNPHIRVGPS					

9/32

Figure 10 (Cont .ued)

Search Analysis for Sequence: MCG4

Search from 1 to 310

Date: September 22,1997

Matrix: pam250 matrix

Score Region from 1 to 310

Maximum possible score: 1598

Aligned sequences:

1. = EST AA074703 phase 1 translation

2. = EST AA134788 phase 3 translation

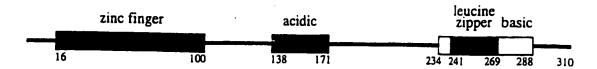
3. = EST AA134788 phase 2 translation

4. = EST AA074703 phase 3 translation

5. = EST AA074703 phase 2 translation

6. = EST AA134788 phase 1 translation

FIGURE 11 Domains of MCG4



 $\mathbf{zinc\ finger\ consensus:\ CX_2HX_4CX_2CX_4HX_2CX_{17}CX_2CX_{18}HX_2CX_{18}CX_2C}$

acidic domain consensus: 9/34 negatively charged amino acids, 0/34 positively charged

basic domain consensus: 13/55 positively charged amino acids, 0/55 negatively charged

leucine zipper domain consensus: LX₆LX₆RX₆LX₆L

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha helix domain: (aa 261) LX₆LXLX₆LXLX₆L (aa 286)

FIGURE 12

FIGURE 12				
			Smallest	•
			Sum Probabili	
		High	•	N
Sequences producing F	ligh-scoring Segment Pairs:	Score	P(N)	N
	reserve more a re	307	3.0e-124	8
gn1 PID e236178	(270752) F25B3.3 (Caenorhabditis ele	202	7.8e-22	5
gi 1293099	(U53884) aimless RasGEF [Dictyosteli	152	3.6e-16	4
gi 1655941	(U67326) Ras-GRF2 [Mus musculus]	150	2.2e-15	3
pir s30356	CDC25 protein homolog - yeast (Candi	150	2.2e-15	3
	CELL DIVISION CONTROL PROTEIN 25	166	2.2e-15 2.6e-15	3
sp P28818 GNRP_RAT	GUANINE NUCLEOTIDE RELEASING PROTEIN		2.6e-15	3
prf 1814463A	guanine nucleotide-releasing factor	166	1.1e-14	1
pir B46199	nucleotide-exchange-factor homolog c	167		3
gn1 PID e238680	(X97560) hypothetical protein L1309	158	3.0e-14	2
pir S22693	CDC25 protein homolog - mouse /gi 50	167	3.7e-14	-
SpiP14771 SC25 YEAST	SCD25 PROTEIN /gi 457494 (M26647) SD	158	4.6e-14	3
SP P26674 STE6_SCHPO	STE6 PROTEIN /pir S28098 ste6 prote	160	5.2e-14	. 2
pir S28407	CDC25 protein homolog - mouse	167	1.2e-13	3
spip27671 GNRP MOUSE	GUANINE NUCLEOTIDE RELEASING PROTEIN	167	1.2e-13	3
gi 1386047	(S62035) Ras-specific guanine nucleo	153	2.0e-13	2
SD 10023421CC25 SACKL	CELL DIVISION CONTROL PROTEIN 25 /pi	142	4.5e~13	2
pir S14177	SCD25 protein - yeast (Saccharomyces	152	5.7e-13	3
gi 433720	(L26584) CDC25 [Homo sapiens]	153	6.0e-13	3
gn1 PID e241744	(268880) T14G10.2 (Caenorhabditis el	157	7.2e-13	1
gi 3484	(X03579) CDC25 protein (aa 1-1588) (136	3.4e-12	3
en I DO 4821 I CC25 VEAST	CELL DIVISION CONTROL PROTEIN 25 /pi	136	3.4e-12	3
gi 915328	(U24070) Muncl3-1 [Rattus norvegicus]	151	5.5e-12	1
pir A46199	nucleotide-exchange-factor homolog c	149	5.6e-12	1
pdb 1PTR	Molecule: Protein Kinase C Delta Ty	136	1.5e-11	1
gi 915330	(U24071) Munc13-2 [Rattus norvegicus]	150	1. 6e-11	2
gi 474982	(D21239) 'C3G protein' [Homo sapiens	131	3.3e-11	3
gi 1763306	(U75361) Muncl3-3 [Rattus norvegicus]	153	6.4e-11	2
	guanine-nucleotide exchange factor C	128	7.8e-11	3
gi 806957	GUANINE NUCLEOTIDE DISSOCIATION STIM	133	1.0e-10	2
pir BVBYL1	LTE1 protein - yeast (Saccharomyces	139	1.9e-10	1
	(D21354) a putative guanine nucleoti	139	2.7e-10	1
gi 452242	LOW TEMPERATURE ESSENTIAL PROTEIN /p	139	2.7e-10	1
	(Z22521) protein kinase C delta (Hom	137	4.0e-10	1
gi 509050	(D10495) protein kinase C delta-type	137	4.6e-10	1
gi 520587	PROTEIN KINASE C. BRAIN ISOZYME (PKC	137	4.7e-10	1
	protein kinase C (EC 2.7.1) delta	137	4.7e-10	1
pir S35704	PROTEIN KINASE C, DELTA TYPE (NPKC-D	137	4.7e-10	1
sp Q05655 KPCU_HUMAN	protein kinase C mu - human /pir A5	137	4.9e-10	1
pir S40279	PROTEIN KINASE C, DELTA TYPE (NPKC-D		9.0e-10	1
sp P09215 KPCD_RAT	(234524) serine/threonine protein ki		1.8e-09	1
gi 520878	(U68142) RalGDS-like [Homo sapiens]	115	3.8e-09	3
gi 1519719	(U68142) KAIGDS-IIKE (NOME SAPIETS)	~~~		

12/32 FIGURE 13(a) (i)

MCG7 - Cloning of a novel human gene that encodes a guanine exchange factor

CGATTTCATTCCTCGCTCCCCACAGGTCCCTCTCCCCAAAATATTCCCATCTTGTCCTAG 60 ISFLAPHRSLSPKYSHLVL 19 CCCATCCCCAGACTATCTCAAGGACCAGCTGTCCCCACGCCCCCGACCTCCACTAGGCC 120 A H P P D Y L K D Q L S P R P P L G TGTGCCACCGCTGCCTGCAGGAAGACGCCCGGTCCCGGGCCGGGTTAGCCCCCATGGGAA 180 LCHPLPAGRRPVPGRVSPMG T Q R L C G R G T Q G W P G S S E Q H V aggaggegacetegteegegggtttgcattetggggtggacgagetggGGGTTCGGTCCG 300 QEATSSAGLHSGVDELGVRS.99 E P G G R L P E R S L G P A H P A P A A 119 TGGCAGGCACCCTGGACCTGGACAAGGGCTGCACGGTGGAGGAGCTGCTCCGCGGGTGCA 420 M A G T L D L D K G C T V E E L L R G C TCGAAGCCTTCGATGACTCCGGGAAGGTGCGGGACCCGCAGCTGGTGCGCATGTTCCTCA 480 I E A F D D S G K V R D P Q L V R M F L TGATGCACCCCTGGTACATCCCCTCCTCTCAGCTGGCGGCCAAGCTGCTCCACATCTACC 540 M M H P W Y I P S S Q L A A K L L · H I Y AACAATCCCGGAAGGACAACTCCAATTCCCTGCAGGTGAAAACGTGCCACCTGGTCAGGT 600 Q Q S R K.D N S N S L Q V K T C H L V R ACTGGATCTCCGCCTTCCCAGCGGAGTTTGACTTGAACCCGGAGTTGGCTGAGCAGATCA 660 Y W I S A F P A E F D L N P E L A E Q I AGGAGCTGAAGGCTCTGCTAGACCAAGAAGGGAACCGACGGCACAGCAGCCCTAATCGACA 720 K E L K A L L D O E G N R R H S S L I D TAGACAGCGTCCCTACCTACAAGTGGAAGCGGCAGGTGACTCAGCGGAACCCTGTGGGAC 780 I D S V P T Y K W K R Q V T Q R N P V G AGAAAAAGCGCAAGATGTCCCTGTTGTTTGACCACCTGGAGCCCATGGAGCTGGCGGAGC 840 Q K K R K M S L L F D H L E P M E L A E ATCTCACCTACTTGGAGTATCGCTCCTTCTGCAAGATCCTGTTTCAGGACTATCACAGTT 900 H L T Y L E Y R S F C K I L F Q D Y H S TCGTGACTCATGGCTGCACTGTGGACAACCCCGTCCTGGAGCGGTTCATCTCCCTCTTCA 960 F V T H G C T V D N P V L E R F I S L F ACAGCGTCTCACAGTGGGTGCAGCTCATGATCCTCAGCAAACCCACAGCCCCGCAGCGGG 1020 NSVSQWVQLMILSKPTAPQR CCCTGGTCATCACACTTTGTCCACGTGGCGGAGAAGCTGCTACAGCTGCAGAACTTCA 1080 A L V I T H F V H V A E K L L Q L Q N F ACACGCTGATGGCAGTGGTCGGGGGCCTGAGCCACAGCTCCATCTCCCGCCTCAAGGAGA 1140 NTLMAVVGGLSHSSISRLKE CCCACAGCCACGTTAGCCCTGAGACCATCAAGCTCTGGGAGGGTCTCACGGAACTAGTGA 1200 THSHVSPETIKLWEGLTELV CGGCGACAGCCAACTATGGCAACTACCGGCGTCGGCTGGCAGCCTGTGTGGGGCTTCCGCT 1260 TATGNYGNYRRRLAACVGFR TCCCGATCCTGGGTGTGCACCTCAAGGACCTGGTGGCCTGCAGCTGGCACTGCCTGACT 1320 FPILGVHLKDLVALQLALPD GGCTGGACCCAGCCCGGACCCGGCTCAACGGGGCCAAGATGAAGCAGCTCTTTAGCATCC 1380 W L D P A R T R L N G A K M K Q L F S I TGGAGGAGCTGGCCATGGTGACCAGCCTGCGGCCACCAGTACAGGCCAACCCCGACCTGC 1440 479 LEELAMVTSLRPPVQANPDL TGAGCCTGCTCACGGTGTCTCTGGATCAGTATCAGACGGAGGATGAGCTGTACCAGCTGT 1500 LSLLTVSLDQYQTEDELYQL CCCTGCAGCGGGAGCCGCGCTCCAAGTCCTCGCCAACCAGCCCCACGAGTTGCACCCCAC 1560 S L Q R E P R S K S S P T S P T S C T P CACCCCGGCCCCCGGTACTGGAGGAGTGGACCTCGGCTGCCAAACCCAAGCTGGATCAGG 1620 PRPPVLEEWTSAAKPKLDQ CCCTCGTGGTGGAGCACATCGAGAAGATGGTGGAGTCTGTGTTCCGGAACTTTGACGTCG 1680

FIGURE 13(a) (ii)

A L V V E H I E K M V E S V F R N F D V ATGGGGATGGCCACATCTCACAGGAAGAATTCCAGATCATCCGTGGGAACTTCCCTTACC 1740 D G D G H I S Q E E F Q I I R G N F P Y TCAGCGCCTTTGGGGACCTCGACCAGAACCAGGATGGCTGCATCAGCAGGGAGGAGATGG 1800 LSAFGDLDQNQDGCISREEM 599 V S Y F L R S S S V L G G R M G F V H N 619 TCCAGGAGAGCAACTCCTTGCGCCCGCCGCCGCCACTGCAAAGCCCTGATCCTGG 1920 PQESNSLRPVACRHCKALIL 639 GCATCTACAAGCAGGGCCTCAAATGCCGAGCCTGTGGAGTGAACTGCCACAAGCAGTGCA 1980 GIYKQGLKCRACGVNCHKQC 659 AGGATCGCCTGTCAGTTGAGTGTCGGCGCAGGGCCCAGAGTGTGAGCCTGGAGGGGTCTG 2040 K D R L S V E C R R R A Q S V S L E G S 679 APSPSPMHSHHHRAFSFSLP 699 GCCCTGGCAGGCGAGGCTCCAGGCCTCCAGAGATCCGTGAGGAGGAGGTACAGACGGTGG 2160 R P G R R G S R P P E I R E E E V Q T V 719 AGGATGGGGTGTTTGACATCCACTTGTAATAGATGCTGTGGTTGGATCAAGGACTCATTC 2220 B D G V F D I H L * \(\sigma\) TGGGGATGGGGTGGGATATGAGGGTGGCATGCAGGCCAGGGCCAGGGCTGGTGT 2340 CCCTAAGGTTGTACAGACTCTTGTGAATATTTGTATTTTCCAGATGGAATAAAAAGGCCC 2400 GTGTAATTAACCTTC(A)n

14/32

FIGURE 13(b)

CGATTTCATTCCTCGCTCCCCACAGGTCCCTCTCCCCAAAATATTCCCATCTTGTCCTAG 60
CCCATCCCCCAGACTATCTCAAGGACCAGCTGTCCCCACGCCCCCGACCTCCACTAGGCC 120
TGTGCCACCCGGTGCTGCAGGAAGACGCCCGGTCCCGGGCCGGGTTAGCCCCATGGGAA 180

* p h g n

* p h g n

CGGGGTTCGGTCCGAGCCCGGTGGGAGGCTCCCGGAGCCTGGGCCCACCC-240

g v r s e p g g r l p e r s l g p a h p

CGCGCCGGCGGCCATGGCAGCCCCTGGACCTGGACCAGGGTGAGGAGGT-360

a p a a M A G T L D L D K G C T V E E L

FIGURE 14

	MAGTLDLDKGCTVEELLRGCIEAFDDSGKVRDPQLVRMFLMMHPW .:.:: . :: : . :: . :: . MSSKVEEDQHQELLTEDQLVARCVECFDVDEEDEVEDIEFVDALFLSHQW	
46	YIPSSQLAAKLLHIYQQSRKDNSNSLQVKTCHLVRYWISAFPAEFDLNPE	95
51	. :::: :. . : : .: . . : : LSDSLSLITHFVNFYQETRNVEQREAVCRAVSFWIEKFPMHFDAQPQ	97
	LAEQIKELKALLDQEGNRHSSLIDIDSVPTYKWKRQVTQRNPVGQKK	143
98	VCAQVVRLKTIAEDINENIRNGL.DVSALPSFAWLRAVSVRNPLAKQTIV	
144	: RKMSLLFDHLEPMELAEHLTYLEYR	
	RVDFETLPTPGTPPPFPIASKKFSLTAFSLSFVQASPSDISTSLSHIDYR	
	SFCKILFQDYHSFVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTAP ::: ::::: ::: : : : : : :::. VLSRISITELKQYVKDGHLRSCPMLERSISVFNNLSNWVQCMILNKTTPK	
219	QRALVITHFVHVAEKLLOLONFNTLMAVVGGLSHSSISRLKETHSHVSPE	268
247	: :: :: : :. : :. : : ERAEILVKFVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTYAVLSND	296
269	TIKLWEGLTELVTATGNYGNYRRRLAAC.VGFRFPILGVHLKDLVALQLA	317
297	IKKELTQLTNLLSAQHNFCEYRKALGACNKKFRIPIIGVHLKDLVAINCS	346
318	LPDWLDPARTRLNGAKMKQLFSILEELAMVTSLRPPV.QANPDLLSLLTV	366
347	GANFEKT. KCISSDKLVKLSKLLSNFLVFNOKGHNLPEMNMDLINTLKV	394
367	SLDQYQTEDELYQLSLQREPRSKSSPTSPTSCTPPPRPPVLEEWTSAAKP	416
	KLDQALVVEHIEKMVESVFRNF <u>DVDGDGHISOEFFQ</u> IIRGNFFYLSAFGE :. .: APDNATVSKHISAMVDAVFKHY <u>DHDRDGFISOEEFQ</u> LIAGNFPFIDAFVN	
		,
	LDONODGCISREEMVSYFLRSS.SVLGGRMGFVHNFOESNSLRPVACRHO : : : : : . : : .	i
	KALILGIYKOGLKCRACGVNCHKOCKDRLSVECRRRAQSVSLEGSAPSP	•
	. :: .: : :. : . : :. : : B NKLLWGILROGFKCKDCGLAVHSCCKSNAVAECRRKSSSNLTRAAEWFA	1
56	6 PMHSHHHRAFSFSLPRPGRRGSRPPEIREEEVQTVEDGVFDIHL 609	
58	: : . : . : .	

FIGURE 15

human	CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG 60)		
human	CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCGACCT CCACTAGGCC 12	20		
human	TOTOCCACCO GOTGOCTICA GGAAGACGCO CGGTCCCGGG CCGGGTTAGC CCCATGGGAA 18	30		
human	CGCAGCGCCT GTGTCGCCCC GGGACTCAAG GCTGGCCTCG CTCAAGTGAA CAGCACGTCC 24	10		
mouse	***tcag** ****ag**** t******* ***a*g***t>			
human	ACCACCCGAC CTCGTCCGCG GGTTTGCATT CTGGGGTGGA CGAGCTGGGG GTTCGGTCCG 30	00		
	acagg	acagg		
mouse	g*****t**a **-*catt** ******** ***aa**aa* g**ct**** **a**aat**>			
human	AGCCCGGTGG GAGGCTCCCG GAGCGCAGCC TGGGCCCCAGC CCACCCCGGG CCGGCGGCCAA 36	50		
mouse	***a*t**** ******tga ***t*t*a*t ****t*t*** ***-*tg**a *****a****>			
human	TOSCASSCAC CCTGGACCTG GACAAGGGCT GCACGGTGGA GGAGCTGCTC CGCGGGTGCA 42	20		
mouse	****ga**** t******* ********** ******** **t**C**t*>			
human	TCGAAGCCTT CGATGACTCC GGGAAGGTGC GGGACCCGCA GCTGGTGCGC ATGTTCCTCA 4	80		
mouse	********** t******** **a**t**a** ***a***** ****t***>			
human	TGATGCACCC CTGGTACATC CCCTCCTCTC AGCTGGCGGC CAAGCTGCTC CACATCTACC 5	40		
mouse	******** ********* **t****** **********			
human	AACAATCCCG GAAGGACAAC TCCAATTCCC TGCAGGTGAA AACGTGCCAC CTGGTCAGGT 6	00		
mouse	*g******* ******* ****************** *****			
human	ACTGGATCTC CGCCTTCCCA GCGGAGTTTG ACTTGAACCC GGAGTTGGCT GAGCAGATCA 6	60		
mouse	********* a******* **a*****C* ******* a***C**** **a******>			
human	AGGAGCTGAA GGCTCTGCTA GACCAAGAAG GGAACCGACG GCACAGCAGC CTAATCGACA 7	20		
mouse	**************************************			
human	TAGACAGCOT . 7	30		
mouse	*c**g**t**			

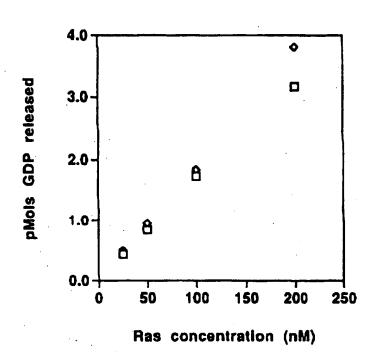
17/32

FIGURE 16

TGGTCGGAAACCGTTACCCGCTCTCCTAGGCCCGGCTAGTGGGGACCCCAACCGCCTGCG 120 * A R L V G T P T A C> GCTGCCCCCCAAGTTCCTCCCTGTTGGCCAGGCATCCAGGTCTCCAGTCTCCGAGCTG 180 G C P S Q V P P C W P G I Q V S S L R A> CGGAGAACCCACCGCCACATGCGGCTGCCCCTTTCCATTCGACCCTGTGGGGAGCCAGGC 240 A E N P P P H A A A P F H S T L W G A R> TTCCGGGGCCCCGTTCCTCTGTGTGAACTGGGCCCCCCGCCCCATTCCCAGACATCAA 300 L P G P R S S C V N W A P R P H S Q T S> GGCCGCGTCTCCAGATAGCCACGATTTCATTCCTCGCTCCCCACAGGTCCCTCTCCCCAA 360 R P R L Q I A T I S P L A P H R S L S P> AATATTCCCATCTTGTCCTAGCCCATCCTCCAGACTATCTCAAGGACCAGCTGTCCCCAC 420 K Y S H L V L A H P P D Y L K D Q L S P> GCCCCGACCTCCACTAGGCCTGTGCCACCCGCTGCCTGCAGGAAGACGCCCGGTCCCGG 480 R P R P P L G L C H P L P A G R R P V P> GCCGGGTTAGCCCCATGGGAACGcagcgcctgtgtggccgcgggactcaaggctggcctg 540 * p h g n G R V S P M G T Q R L C G R G T Q G W P> gctcaagtgaacagcacgtccaggaggcgacctcgtccgcgggtttgcattctggggtgg 600 G S E Q H V Q E A T S S A G L H S G V> acgagctggGGGTTCGGTCCGAGCCCGGTGGGAGGCTCCCGGAGCGCAGCCTGGGCCCAG 660 DELGVRSEPGGRLPERSLGP> CCCACCCGCGCGGCGGCCATGGCAGGCACCCTGGACCTGGACAAGGGCTGCACGGTGG 720 A H P A P A A M A G T L D L D K G C T V>

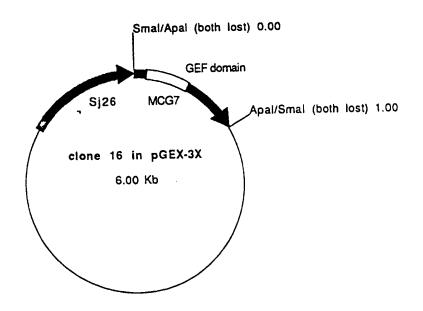
18/32

FIGURE 17



WO 98/53061 PCT/AU98/00380

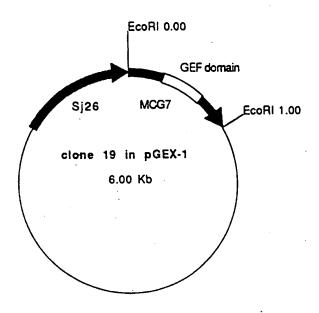
19/32
FIGURE 18 (Cont. I)



Plasmid name: clone 16 in pGEX-3X

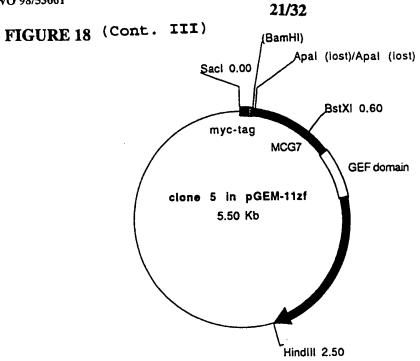
Plasmid size: 6.00 kb

FIGURE 18 (Cont. II)



Plasmid name: clone 19 in pGEX-1

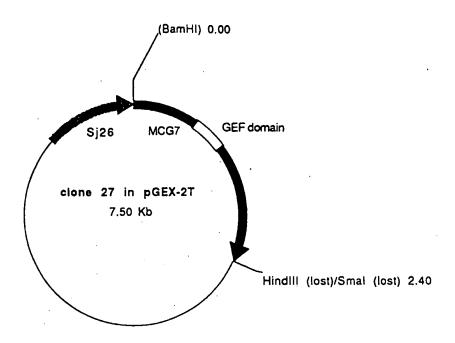
Plasmid size: 6.00 kb



Plasmid name: clone 5 in pGEM-11zf

Plasmid size: 5.50 kb

FIGURE 18 (Cont. IV)



Plasmid name: clone 27 in pGEX-2T

Plasmid size: 7.50 kb

23/32

FIGURE 19

GCCC	GCCG	CC A	TG C et P	CG C	CC T'	TA C eu L	TG C eu P 5	CC C	TG C eu A	GC C rg L	eu c	GC C ys A 10	rg L	rg r eu T	rp GG	49
CCC Pro	CGC Arg 15	AAC Asn	CCT Pro	CCC '	TCC (Ser)	CGG Arg	CTC	CTC Leu	GGA Gly	GCG (GCC Ala 25	GCC Ala	GGG Gly	CAG Gln	CGG Arg	97
TCC Ser 30	AGA Arg	CCC Pro	AGT : Ser	ACT Thr	TAT ' Tyr '	TAT (GAA Glu	CTG Leu	TTG Leu	GGG (Gly 40	GTG Val	CAT His	CCT Pro	GGT Gly	GCC Ala 45	145
AGC Ser	ACT Thr	GAG Glu	GAA Glu	GTT Val	AAA Lys	CGA (GCT Ala	TTC Phe	TTC Phe 55	TCC . Ser	AAG Lys	TCC Ser	AAA Lys	GAG Glu 60	CTG Leu	193
CAC His	CCA Pro	GAC Asp	CGG Arg 65	GAC Asp	CCT (GGG . Gly .	AAC Asn	CCA Pro 70	AGC Ser	CTG Leu	CAC His	AGC Ser	CGC Arg 75	TTT Phe	GTG Val	241
GAG Glu	CTG Leu	AGC Ser 80	GAG Glu	GCA Ala	TAC Tyr	CGT Arg	GTG Val 85	CTC Leu	AGC Ser	CGT	GAG Glu	CAG Gln 90	AGC Ser	CGC Arg	Arg CGC	289
AGC Ser	TAT Tyr 95	GAT Asp	GAC Asp	CAG Gln	Leu	CGC Arg 100	TCA Ser	GGT Gly	AGT Ser	CCC Pro	CCA Pro 105	AAG Lys	TCT Ser	CCA Pro	Arg	337
ACC Thr 110	ACA Thr	GTC Val	CAT His	GAC Asp	AAG Lys 115	TCT Ser	GCC Ala	CAC His	CAA Gln	ACA Thr 120	CAC His	AGC Ser	TCC Ser	TGG Trp	ACA Thr 125	385
CCC	CCC Pro	AAC Asn	GCA Ala	CAG Gln 130	TAC Tyr	TGG Trp	TCC Ser	CAG Gln	TTT Phe 135	CAC His	AGC Ser	GTG Val	AGG Arg	CCA Pro 140	CAG Gln	433
GGG Gly	CCC	CAG Gln	TTG Leu 145	AGG Arg	CAG Gln	CAG Gln	CAA Gln	CAC His 150	AAA Lys	CAA Gln	AAC Asn	ъys	CAA Gln 155	GTG Val	CTG Leu	481
GGG	TAC	TGC Cys 160	Leu	CTC Leu	CTC Leu	ATG Met	CTG Leu 165	Ala	GGC Gly	ATG Met	GGC	CTG Leu 170	*****	TAC Tyr	ATT Ile	529
GCC Ala	TTC Phe 175	Arg	AAG Lys	GTG Val	AAG Lys	CAG Gln 180	Met	CAC His	CTT Leu	AAC Asn	TTC Phe 185	- MCC	GAT Asp	GAM Glu	AAG Lys	577
CAT		2 ATC	- ATC	: ACA	GCC	TTC	TAC	: AAC	GAF	GCC	CGG	G GCI	A CG(GC	CAGG	625

WO 98/53061 FIGURE 19 (cont: led	24/32	PCT/AU98/00380
Asp Arg Ile Ile Thr A	la Phe Tyr Asn Glu Ala Arg Ala	Arg Ala Arg
190 l	95 200	205
GCC AAC AGA GGC ATC C	TT CAG CAG GAG CGA CAA CGG CTA	GGG CAG CGG 673
Ala 7.:n Arg Gly Ile L	eu Gln Gln Glu Arg Gln Arg Leu	Gly Gln Arg
210	215	220
CAG CCG CCA CCA TCC G	AG CCA ACC CAA GGC CCC GAG ATC	GTG CCC CGG 721
Gln Pro Pro Pro Ser G	lu Pro Thr Gln Gly Pro Glu Ile	Val Pro Arg
225	230	235
GGC GCC GGC CCC TGA GC Gly Ala Gly Pro * 240	EGGCTC ACCTGGATGG GGCCTGCAGT G	CGTTCCCGC 773
TTTGCTTCCT TCCCTGGACG	GCCCGCTCCC CGAAACGCGC GCAATAA	AGT GATTCGCAG 832

>sp|P08622|DNAJ_ECOLI DNAJ PROTEIN >pir||HHECDJ heat shock protein dnaJ -Escherichia coli >gi|145769 (M12565) heat shock protein dnaJ [Escherichia coli] >gi|216441 (D10483) dnaJ protein [Escherichia colil

Length = 376

Score = 138 (63.7 bits), Expect = 1.2e-10, P = 1.2e-10 Identities = 25/62 (40%), Positives = 39/62 (62%)

35 YYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRS 94 Query:

YYE+LGV A E+++A+ + + HPDR+ G+ ++F E+ EAY VL+ Q R + 6 YYEILGVSKTAEEREIRKAYKRLAMKYHPDRNQGDKEAEAKFKEIKEAYEVLTDSQKRAA 65 Sbjct:

95 YD 96 Query:

YD

66 YD 67 Sbjct:

Score = 98 (45.2 bits), Expect = 5.2e-12. Sum P(3) = 5.2e-12 Identities = 17/37 (45%), Positives = 28/37 (75%)

Query: 28 QRSRPSTYYELLGVHPGASTEEVKRAFFSKSKELHPD 64 ++ R T+YE+LGV A+ E+K AF+++SK++HPD Sbjct: 22 KKIRQRTHYEVLGVESTATLSEIKSAFYAQSKKVHPD 58

Score = 74 (34.1 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12 Identities = 17/32 (53%), Positives = 19/32 (59%)

Query: 71 SLHSRFVELSEAYRVLSREQSRRSYDDQLRSG 102 S + F+EL AY VL R RR YD QLR G Sbjct: 64 SATASFLELKNAYDVLRRPADRRLYDYQLRGG 95

Score = 39 (18.0 bits), Expect = 5.2e-12, Sum P(3) = 5.2e-12 Identities = 10/42 (23%), Positives = 19/42 (45%)

Query: 162 LLMLAGMGLHYIAFRKVKQMHLNFMDEKDRIITAFYNEARAR 203 L+++AG Y+ Q L++++D I F + R Sbjct: 158 LVLVAGYNGGYLYLLAYNQKQLDKLIDEDEIAKCFLRQKEFR 199

>gnl|PID|e281266 (Z81030) C01G10.12 (Caenorhabditis elegans)
Length = 191

Score = 96 (44.3 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09 :Identities = 17/41 (41%), Positives = 27/41 (65%)

Query: 35 YYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSR 75 YYE++GV A+ +E++ AF K+K+LHPD+ + SR Sbjct: 19 YYELIGVSASATRQEIRDAFLKKTKQLHPDQSRKSSKSDSR 59

Score = 54 (24.9 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09 Identities = 10/22 (45%), Positives = 15/22 (68%)

Query: 75 RFVZLSEAYRVLSREQSPFSYD 96 +F+ + EAY VL E+ R+ YD Sbjct: 71 QFMLVKEAYDVLRMEKRKEYD 92

Score = 35 (16.1 bits), Expect = 1.8e-09, Sum P(3) = 1.8e-09Identities = 9/44 (20%), Positives = 22/44 (50%)

Query: 141 QGPQLRQQQHKQNKQVLGYCLLIMLAGMGLHYIAFRKVKQMHLN 184 + P+ + KQ ++L ++A +G + + RK++ L+ Sbjct: 145 RNPEDEYLREKQKNRMLVVLAATVMALIGANIVYIRKLQADRLS 188

>sp|Q10209|YAY1_SCHPO HYPOTHETICAL 44.8 KD PROTEIN C4H3.01 IN CHROMOSOME I >gi|1184014 (Z69380) unknown [Schizosaccharomyces pombe] Length = 392

Score = 84 (38.8 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08Identities = 13/35 (36%), Positives = 25/36 (69%)

35 YYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNP 70 Ouerv: YY+LLG+ A+ ++K+A+ + + HPD++P +P 9 YYDLLGISTDATAVDIKKAYRKLAVKYHPDKNPDDP 44

Sbjct:

Score = 64 (29.5 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08Identities = 14/40 (35%), Positives = 23/40 (57%)

75 REVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPPTTVHD 114 Ouery: +F ++SEAY+VL E+ R YD + + P+ T +D 50 KFQKISEAYQVLGDEKLRSQYDQFGKEKAVPEQGFTDAYD 89 Sbjct:

Score = 37 (17.1 bits), Expect = 4.1e-08, Sum P(3) = 4.1e-08Identities = 9/29 (31%), Positives = 15/29 (51%)

190 DRIITAFYNEARARANRGILQQERQRL 218 Query: DR A E A A+ + +++ RQR+ 149 DRKKNAQIREREALAKREQEMIEDRRQRI 177 Sbict:

Score = 33 (15.2 bits). Expect = 0.00081, Sum P(3) = 0.00081Identities = 8/19 (42%), Positives = 11/19 (57%)

140 PQGPQLRQQQHKQNKQVLG 158 Ouerv: PQG + Q+ + QVLG 44 PQGASEKFQKISEAYQVLG 62 Sbjct:

FIGURE 23

>gnl|PID|e253406 (X77635) tumorous imaginal discs [Drosophila virilis] >gnl|PID|e263866 (Y07700) Tid58 protein [Drosophila virilis] Length = 529

Score = 153 (70.6 bits), Expect = 9.7e-13, P = 9.7e-13Identities = 27/71 (38%), Positives = 44/71 (61%)

26 AGQRSRPSTYYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRV 85 Query: + R + YY LGV A+ +++K+A++ +K+ HPD + +P +F ++SEAY V 72 SSSRMQAKDYYATLGVAKNANAKDIKKAYYELAKKYHPDINKDDPDASKKFQUVSEAYEV 131 Sbjct:

86 LSREQSRRSYD 96 Query: LS +Q RR YD

132 LSDDQKRREYD 142 Sbict:

28/32

MCG18 HDJ-2 HDJ-1 HSJ1	MVKETTYYDVLGVKPNATQEELKKAYRKLALKYHPDKNPNEGEXFKQISQAYEV MGKDYYQTLGLARGASDEEIKRAYRRQALRYHPDKNKEPGAEEKFKEIAEAYDV M-ASYYEILDVPRSASADDIKKAYRRKALQWHPDKNPDNKEFAEKKFKEVAEAYEV
MCG18 HDJ-2 HDJ-1 HSJ1	AGQRSRPSTYYELLGVHST-EEVKRAFFS LSDAKKRELYDKGGEQAIKEGGAGGG
MCG18 HDJ-2 HDJ-1 HSJ1	KSKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRT GRMQRERRGKNVVHQLSVTLEDLYNGATRKLALQKNVICDKCEGRGGKKGAVECCPNCRG GRNPFDTFFGQRNGEECMDIDDPFSGFPMCMGGFTNVNFGRSRSAQEPARKKQDPPVT SGDPFAELFDDLGPFSELQNRGSRHSGPFFTFSSSFPGHSDFSSSSFSFSPGAGAFRS
MCG18 HDJ-2 HDJ-1 HSJ1	TVHDKSAHOTHSSWTPPNAQYWSQFHSVRPQGPQLRQQQHKQN TCMQIRIHQIGPCMVQQIQSVCMECQGHGERISPK-DRCKSCNGRKIVREKKILEVHIDK HDLRVSLEEIYSGCTKCMKISH-KRLNPDGKSIRNEDKILTIEVKK VSTSTTFVQGRRITTRRIMENGQ-ERVEVEEDGQLKSVTINGVPD
MCG18 HDJ-2 HDJ-1 HSJ1	KQVLGYCLLLMLAGMGLHYIAFRKVKQMHLNFMDE-KDRIITAFYNEARARARAN GMKDGQKITFHGEGDQEPGLEPGDIIIVLDQKDHAVFTRRGEDLFMCMDIQLVEALCGFQ GWKEGTKITFPKEGDQTSNNIPADIVFVLKDKPHNIFKRDGSDVIYPARISLREALCGCT DLARGLELSR-REQQP-SVTSRSGGTQVQQTPASCPLD-SDLSEDEDLQLAMAYSLSE
MCG18 HDJ-2 HDJ-1 HSJ1	RGILQQERQRLCQRQPP-PSEPTQGPEIVPRGAGP KPISTLINRTIVITSHPGQIVKHGDIKCVLNEGMPIYRRPYEKGRLIIEFKVNFPENGFL VNVPTLLDGRTIPVVFKDVIRPGMRRKVPGEGLPLPKTPEKRGDLIIEFEVIFPERI MEAAGKKPAGGREAQHR-RQGRPRPSTKIQAWGGPRRVRGVKQPNAVHPQR-RR
MCG18 HDJ-2 HDJ-1 HSJ1	SPDKLSLLEKLLPERKEVEETDEMDQVELVDFDPNQERRRHYNGEAYEDDEHHPRGGVQC PQTSRTVLEQVLPILIQILTGGSDSLWEEKRGVS
MCG18 HDJ-2 HDJ-1 HSJ1	QTS

⁼ amino acid identity in all 4 proteins

^{. =} conservative substitution

~	-C-2	~~	יתרים	NG C (יתיכר	ירכס	TCC	TCC	TCA	TGC	CGI	ccc	TGT	r T G(CTCC	CAGO	CTG	CCC	CTGC	60
CAA	(GGA	iGC (.10	, GC	. 100			,		M	P	s	L	L	L	Q	L	P	L	10
cct	ጉጥልባ	YGC (rgge	TG'	rgg(CCGC	ATA	\GC(TT	CCA	TCC	GA	CTT(CTC	ACAG	3CC(GCC	ACA	GGGC	120
D .	L	ر د	R	L	w	P	н	s	L	s	I	R	L	L	T	A	A	T	G	30
v.	ים ים	י	TC(CCT	ACT	AATT	rac:	YTAT	GAAT	TGI	TGC	3GC(GTG	CAT	CCG	GGT	GCC.	AGC	GCTG	180
Λ Ο	R	s	v	P	т	N	Y	Y	E	L	L	G	v	н	P	G	A	S	A	50
•					ماليات	- Arter	יייין	ACC	AAG'	rca.	LAA(GAG	CTA	CAC	CCT	GAT	CGA	GAC	CCTG	240
AA	E E	*111	K K	R	A	F	F	T	К	S	ĸ	E	L	н	P	D	R	ם	P	70
-		יה מים	تدرد. ••	ርጥር.	ርልጥ	AGC	cgc'	rrr	GTG	GAGO	TG	ТАА	GAG	GCA	TAT	CGA	GTG	CTC	AGTC	300
~	N	יתט	A C	T.	н	s	R	F	v	E	L	N	E	Ā	Y	R	٧	L	s	90
	14		 	CCM	 	2 2 C	ጥልጥ	CAC	CAC	CAGO	TG	CAT	TCA	GCC	AGT	CCT	CCA	AAG	TCTT	360
			AGT	-	-	AAC	V.	חייים	u	Q	Τ.	н	s	A	s	P	P	K	· S	110
R	E		S			N		~~ ~~~	 NCC	~ X X (~AG	ACA	CAC	AGC	'AGC	TCC	TGG	GAP	rcccc	420
CA	GGG.	AGC	ACA	.GCC	GAG	CCI	AAG	TAT	ACG	CAA	-AG	n m	ייייייייייייייייייייייייייייייייייייייי		S	s	W	E	P	130
S	G	S	T	A	E	P	K	Y	T		_ 	1			ccc	CCC	···	ייירי	AGGA	480
CC	AAC	GCI	CAA	TAC	TGG	GCC	CAG	TTC	CAC	AGT	GTG	AGC	-	CAC	-	-	-		AAGGA	150
P	N	A	Q	. Y	W	A	Q	F	Н	S	V	R	P	Q					~~~	
AG	CAG	CAC	CGI	LAA	CAC	CAAC	CAC	CGG	GTC	CTG	GGG	TAC	TGC	CT	CTC	3CT(ZA'IX	3G1V	GCAG	170
ĸ	Q	Q	R	ĸ	H	N	Q	R	V	L	G	Y	С	L	L	L	M	V	A	170
GC	ATC	GGG	CTC	GCA	CTA!	rgti	rgco	TTC	CAGG	AAG	CTC	GAG	GCA(GTY	GCA'	rcg	CAG	CTT	CATGG	, 600
G	М	G	L	н	Y	v	A	F	R	K	L	E	Q	٧	Н	R	S	F	M	190
Δr	rca i	AAA	GGA	CCG	GAT	CAT	raci	AGC	CATC	CTAC	'AA'	rga:	CAC'	TCG	GGC	CAG	GGC	CAG	GGCCA	4 660
ח	E	к	D	R	r	I	T	A	I	Y	14	D	T	R	A	R	. A	R	. А	210
- A	CAG:	\GC	CAG	GAT	TCA	GCA!	CGA	GCG	CCA	CGAC	GAG	GCA	.GCA	.GCC	TCG	GGC	AGA	ACC	CTCC	2 720
N	R	A	R	I	Q	Q	E	R	Н	E	R	Q	2	F	P	L A	. E	E	S	230
т Т	כככי	TCC	AGA	AAG	CTC	CAG	GAT	CAT	GCC	CCA	GGA	CAC	AAC	CCC	CTC	BAGA	\GG(TT	ACTA	A 780
	P																			245
1	W	- -	יייייייייייייייייייייייייייייייייייייי	י ב	رتاماتار 	יייריר	ייטיי	יככי	TGC	TGC	CTC	TC	CAG	AAC'	raci	ACG'	rgc.	AAT.	AAACI	C 840
																				849
7	TTT	LC:	\G (ı)n																

PCT/AU98/00380

30/32

FIGURE 26

human	MCG18	MPPLLPLRICRLWPRNPPSRLIGAAAGQRSRPSTYYELLGVHPGASTEEVRRAFFSK
mouse	MCG18	MPSLLLQLPLRLCRLWPHSLSIRLLTAATGQRSVPTNYYELLGVHPGASAEEIKRAFFTK
	MCG18	SKELHPDRDPCNPSLHSRFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRTTVHDKSA SKELHPDRDPCNPALHSRFVELNEAYRVLSREESRRNYDHQLHSASPPKSSGSTAEPKYT
mouse	MCG18	*******************************
human	MCG18	HQTHSS-WTPPNAQYWSQFHSVRPQGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFR
mouse	MCG18	QQTHSSSWEPPNAQYWAQFHSVRPQGPESRKQQRKHNQRVLGYCLLLMVAGMGLHYVAFR
human	MCG18	KVKQMHLNFMDEKDRIITAFYNEARARARANRGILQQERQRLGQRQPPPSEPTQGPE
mouse	MCG18	KLEQVHRSFMDEKDRIITAIYNDTRARARANRARIQQERHERQQPRAEPSLPPESSR
human	MCG18	IVPRGAGP
mouse	MCG18	IMPQDTSP
		*

31/32

FIGURE 27

ttgaagtctagccccatcctggtccaatgcgctcttggtagcctcctrtcccagctgccc 60
* S L A P S W S N A L L V A S F P S C P

geegeegecATGCCGCCCTTACTGCCCCTGCGCCTGTGCCGGCTGTGGCCCCGCAACCC 120
 P A A M P P L L P L R L C R L W P R N P>

WO 98/53061 PCT/AU98/00380

32/32

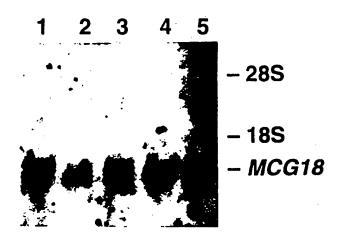


FIGURE 28

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 98/00380

		PCIIAU	78/00380
A.	CLASSIFICATION OF SUBJECT MATTER		
Int Cl ⁶ :	C12N 15/12; C07K 14/47; C07K 16/18; G01N 33/5	3	·
According to	International Patent Classification (IPC) or to both na	tional classification and IPC	
В.	FIELDS SEARCHED		
	umentation searched (classification system followed by class	sification symbols)	
I/C:	WPAT (D gene) Sequences provided by Applica		
Documentatio	on searched other than minimum documentation to the extent	that such documents are included in the	ne fields searched
Electronic dat :EMBL, Ge	ta base consulted during the international search (name of deenebank, Swiss Prot and PIR: Sequences provided :	ata base and, where practicable, search by applicant	terms used)
C.	DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appro	opriate, of the relevant passages	Relevant to claim No.
P,X	Kedra D, Seroussi E, Fransson I, Trifunovic J, Blennow E, Mehlin H, Dumanski J, Human Ge 611-619 The germinal centre kinase gene and a located in the vicinity of the PYGM gene on 11 EMBL AC Y12339	Clark M, Lagercranz J, enetics, October 1997 100(5-6) a novel CDC25-like gene are	1-3,8-10,15-18
P,X	Guru S C, Agarwal S K, Manickain P, Olufen July 1997 7(7) 725-735. A transcript map for the multiple endocrine neoplasia type I locus TREMBL AC 014616	ni S E, et al Genome Research, the 2.8-Mb region containing	1. 4-5, 8, 11-12, 15, 19-21
X	Further documents are listed in the continuation of Box C	See patent family a	nnex
"A" do no no "E" ea in "L" do or ar "O" do eo	recial categories of cited documents: "T" comment defining the general state of the art which is of considered to be of particular relevance which is determined to the published on or after the ternational filing date occument which may throw doubts on priority claim(s) which is cited to establish the publication date of mother citation or other special reason (as specified) occument referring to an oral disclosure, use, whibition or other means occument published prior to the international filing ate but later than the priority date claimed	be considered novel or cannot be conventive step when the document document of particular relevance; be considered to involve an inventicular combined with one or more other step to the combination being obvious to a pe	h the application but cited to inderlying the invention he claimed invention cannot onsidered to involve an is taken alone he claimed invention cannot ive step when the document is such documents, such rson skilled in the art
	e actual completion of the international search	Date of mailing of the international set 2 0 JUL 199	
Name and AUSTRAI PO BOX : WODEN AUSTRA	mailing address of the ISA/AU LIAN PATENT OFFICE 200 ACT 2606	Authorized officer GILLIAN ALLEN Telephone No.: (02) 6283 2266	; 7,7(1c)

INTERNATIONAL SEARCH REPORT

international Application No.

PCT/AU 98/00380

Box 1	Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This Inte	ernational Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following
1.	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	X Claims Nos.: 1, 2, 4, 6 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
	They are to known groups of proteins and lack distinguishing features which would enable a meaningful search.
3.	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)
Вох П	Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
Inver zinc Inver whic Inver	ernational Searching Authority found multiple inventions in this international application, as follows: Intion 1, defined by claims 2, 3, 9, 10, 16-18, is to nucleotide sequences, amino acid sequences and proteins with a finger domain. Intion 2, defined by claims 4, 5, 11, 12, 19-21, is to nucleotide sequences and amino acid sequences and proteins that are guanine exchange factors. Intion 3, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins that are heat shock proteins or heat shock binding proteins.
1.	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2.	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remar	k on Protest
	No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 98/00380

(Continuat	tion) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EMBL AC AF012106 DT 6 November 1997 Lloyd S E and Thakker R V DE Homo Sapiens DnaJ protein (HSPF ₂)mRNA, complete cds	1,6-8,13- 15,22-24
P,X	EMBL AC AF 036875 DT 20 May 1998 Silins G, Grimmond S, Hayward N DE Mus musculus multiple endocrine neoplasia type I candidate protein number 18 mRNA, complete cds	1,6-8,13- 15,22-24
·		
ŧ.		



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Paten	t Classification 6:		(11) International Publication Number	r: WO 98/53061
C12N 15/12, C0 33/53	7K 14/47, 16/18, G01N	A1	(43	i) International Publication Date:	26 November 1998 (26.11.98)
(21) International Appli	ication Number: PCT/AU	98/003	80	(74) Agents: HUGHES, E., John, L. Little Collins Street, Melbou	
(22) International Filing	Date: 22 May 1998 (2)	22.05.9	8)		
CIL OF THE QU RESEARCH [AU 4029 (AU). (72) Inventors; and (75) Inventors/Applicant [AU/AU]; 13 Pr SILINS, Ginters QLD 4061 (AU) Research Counci (GB). GARTSID Camp Hill, QLD	23 May 1997 (23.05.97) 23 May 1997 (23.05.97) 23 May 1997 (23.05.97) 22 January 1998 (22.01.98) 22 January 1998 (22.01.98) 22 January 1998 (22.01.98) 22 January 1998 (22.01.98) 23 January 1998 (22.01.98) 24 January 1998 (22.01.98) 25 January 1998 (22.01.98) 26 January 1998 (22.01.98) 26 January 1998 (22.01.98) 27 January 1998 (22.01.98) 28 January 1998 (22.01.98) 29 January 1998 (22.01.98) 29 January 1998 (22.01.98) 20 January 1998 (22.01.98) 20 January 1998 (22.01.98) 21 January 1998 (22.01.98) 22 January 1998 (22.01.98) 23 January 1998 (22.01.98) 24 January 1998 (22.01.98) 25 January 1998 (22.01.98) 26 January 1998 (22.01.98) 26 January 1998 (22.01.98) 27 January 1998 (22.01.98) 28 January 1998 (22.01.98) 29 January 1998 (22.01.98) 29 January 1998 (22.01.98) 20 January 1998 (22.01.98) 20 January 1998 (22.01.98) 20 January 1998 (22.01.98) 21 January 1998 (22.01.98) 22 January 1998 (22.01.98) 23 January 1998 (22.01.98) 24 January 1998 (22.01.98) 25 January 1998 (22.01.98) 26 January 1998 (22.01.98) 26 January 1998 (22.01.98) 27 January 1998 (22.01.98) 27 January 1998 (22.01.98) 28 January 1998 (22.01.98) 29 January 1998 (22.01.98) 29 January 1998 (22.01.98) 20 January 1998 (22.01.98) 21 January 1998 (22.01.98) 22 January 1998 (22.01.98) 23 January 1998 (22.01.98) 24 January 1998 (22.01.98) 25 January 1998 (22.01.98) 26 January 1998 (22.01.98) 26 January 1998 (22.01.98) 27 January 1998 (22.01.98) 28 January 1998 (22.01.98) 28 January 1998 (22.01.98) 29 January 1998 (22.01.98) 20 Jan	A A A A A A A A A A A A A A A A A A A	UUUUUUU N-LLD aas J). ap, cal D et,	GH, GM, GW, HU, ID, IL, LC, LK, LR, LS, LT, LU, I MX, NO, NZ, PL, PT, RO, TJ, TM, TR, TT, UA, UG, I patent (GH, GM, KE, LS, M patent (AM, AZ, BY, KG, K patent (AT, BE, CH, CY, I	DE, DK, EE, ES, FI, GB, GE, IS, JP, KE, KG, KP, KR, KZ, LV, MD, MG, MK, MN, MW, RU, SD, SE, SG, SI, SK, SL, US, UZ, VN, YU, ZW, ARIPO IW, SD, SZ, UG, ZW), Eurasian Z, MD, RU, TJ, TM), European DE, DK, ES, FI, FR, GB, GR, SE), OAPI patent (BF, BJ, CF, MR, NE, SN, TD, TG).

(54) Title: THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN

(57) Abstract

The present invention relates generally to three novel human genes with gene regulatory function. These genes encode a zinc finger protein, a guanine nucleotide exchange protein and a heat shock protein or heat shock binding protein. The invention includes derivatives and mammalian animal, insect, nematodes, avian and microbial homologues of these genes. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL AM AT AU AZ BA BB BE BF BG BJ BR BY CA CF CG CH CI CM CN CU DE DK EE	Albania Armenia Austria Australia Azerbaijan Bosnia and Herzegovina Barbados Belgium Burkina Faso Bulgaria Benin Brazil Belarus Canada Central African Republic Congo Switzerland Côte d'Ivoire Cameroon China Cuba Czech Republic Germany Denmark Estonia	ES FI FR GA GB GC GH GN GR HU IE IL IS IT JP KE KG KP LC LI LK LR	Spain Finland France Gabon United Kingdom Georgia Ghana Guinea Greece Hungary Ireland Israel Iceland Italy Japan Kenya Kyrgyzstan Democratic People's Republic of Korea Republic of Korea Kazakstan Saint Lucia Liechtenstein Sri Lanka Liberia	LS LT LU LV MC MD MG MK MN MN MN MN MN NE NL NO NZ PL PT RO RU SD SE SG	Lesotho Lithuania Luxembourg Latvia Monaco Republic of Moldova Madagascar The former Yugoslav Republic of Macedonia Mali Mongolia Mauritania Malawi Mexico Niger Netherlands Norway New Zealand Poland Portugal Romania Russian Federation Sudan Sweden Singapore	SI SK SN SZ TD TG TJ TM TR TT UA UG US UZ VN YU ZW	Slovenia Slovakia Senegal Swaziland Chad Togo Tajikistan Turkmenistan Turkey Trinidad and Tobago Ukraine Uganda United States of America Uzbekistan Viet Nam Yugoslavia Zimbabwe	
-------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

WO 98/53061 PCT/AU98/00380

- 1 -

THREE NOVEL GENES ENCODING A ZINC FINGER PROTEIN, A GUANINE, NUCLEOTIDE EXCHANGE FACTOR AND A HEAT SHOCK PROTEIN OR HEAT SHOCK BINDING PROTEIN

FIELD OF THE INVENTION

5 The present invention relates generally to a novel human gene and its derivatives and to mammalian, animal, insect, nematodes, avian and microbial homologues thereof. The present invention further provides pharmaceutical compositions and diagnostic agents as well as genetic molecules useful in gene replacement therapy and recombinant molecules useful in protein replacement therapy.

10

BACKGROUND OF THE INVENTION

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

15

The increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. There is growing need to develop recombinant and genetic molecules for use in diagnosis and in conventional pharmaceutical preparations as well as in gene and protein replacement therapies.

20

In work leading up to the present invention, the inventors sought to identify and clone human genes which might be useful as potential diagnostic and/or therapeutic agents. Molecules of particular interest targeted by the inventors were gene regulators including regulatory proteins, signal transducters and heat shock proteins.

25

Gene expression generally requires interaction between a regulatory protein and an appropriate recognition sequence of a target gene. Regulatory proteins comprise in many cases a domain or motif which facilitates binding to DNA. One particular motif comprises small sequence units repeated in tandem with each unit folded about a zinc atom to form separate structural domains.

30 This motif is now referred to as a zinc finger domain. Such a domain is generally defined by the number of cysteine (C) and histidine (H) residues.

In addition, knowledge of cellular interaction in the control of cell proliferation is essential in the rational design of specific therapeutic strategies aimed at controlling proliferative disorders. Such proliferative disorders including a range of cancers, inflammatory conditions and atherosclerosis. An important aspect of cellular interaction is in signal transduction via receptors to intracellular transducers. One key signal transducer is Ras which couples the receptors for diverse extracellular signals to different effectors. Ras directly activates the downstream kinase Raf which in turn induces the mitogen activated protein kinase (MAPK) cascade.

Another regulatory mechanism involves heat shock proteins. The *Escherichia coli* heat shock protein, DnaJ, is the founding member of a family of proteins which are associated with protein folding, protein complex assembly and transit through subcellular components.

Prokaryotic and eukaryotic DnaJ homologues have a modular organisation consisting of a J domain, a glycine-rich spacer, CXXCXGXG [SEQ ID NO:1] repeats and a C-terminal region with no obvious sequence features, as well as additional sequences for protein targeting. The J domain is anticipated to mediate interaction with heat shock 70 proteins (Hsp70) and consists of some 70 amino acids, frequently located at the N-terminus of the protein.

In accordance with the present invention, a genes have been identified from the human genome which encodes proteins having a regulatory role. One gene, in accordance with the present invention encodes a protein with an N-terminal region resembling a zinc-finger domain of a novel type. Another gene encodes a protein involved in guanine nucleotide exchange factor (GEF) signalling pathways. Yet another gene encodes a protein which is a heat shock protein or heat shock-like protein which may have a role in tumour suppression.

25

SUMMARY OF THE INVENTION

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Sequence identity numbers (SEQ ID NOs.) for nucleotide and amino acid sequences referred to in the subject specification are defined after the bibliography. A summary of SEQ ID NOs. is also given in Table 1.

- 5 One aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
- 10 Another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an (HC₃)₂ type.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:2 defines the gene, mcg4. This gene encodes a product, MCG4, having an amino acid sequence set forth in SEQ ID NO:3.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg4 gene portion, which mcg4 gene portion is capable of encoding an MCG4 polypeptide or a 30 functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg4 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.

A further aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

Yet another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

25

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions to the

nucleotide sequence set forth in (i), (ii) or (iii).

The nucleotide sequence set forth in SEQ ID NO:4 or 6 defines the gene, mcg7. This gene encodes a product, MCG7, having an amino acid sequence set forth in SEQ ID NO:5 or 7.

5

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg7 gene portion, which mcg7 gene portion is capable of encoding an MCG7 polypeptide or a functional or immunologically interactive derivative thereof.

10

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg7 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

Another aspect of the present invention contemplates a method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

Yet another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock binding protein or a derivative thereof.

Another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 5 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
 - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 41°C to the nucleotide sequence set forth in (i), (ii) or (iii).

10

The nucleotide sequence set forth in SEQ ID NO:8 defines the gene, mcg18. This gene encodes a product, MCG18, having an amino acid sequence set forth in SEQ ID NO:7.

Even yet another aspect of the present invention provides a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg18 gene portion, which mcg18 gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Still yet another aspect of the present invention contemplates a method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg18 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

25

Even still a further aspect of the present invention relates to a method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

30

Another aspect of the present invention contemplates a method for detecting MCG18 or a

PCT/AU98/00380

derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

- 5

A summary of SEQ ID Nos. referred to in the subject specification is shown in Table 1.

TABLE 1
SUMMARY OF SEQ ID Nos.

5	SEQ ID NO.	DESCRIPTION
	1	amino acid repeat sequence in DnaJ homologues
	2	Nucleotide sequence of mcg4
	3	amino acid sequence of MCG4
	4	nucleotide sequence of mcg7
10	5	amino acid sequence of MCG7
	6	nucleotide sequence of mcg7 within exon of
		nucleotides 183-288
	7	amino acid sequence of MCG7 within exon of
		nucleotide 183-288
	8	nucleotide sequence of mcg18
	9	amino acid sequence of MCG18
15	10-18	amino acid sequence identified using BESTFIT
	19	sequence of pGEX and mcg7 junction
	20	sequence of pGEX and mcg7 junction
	21	nucleotide sequence of myc-tag/mcg7 junction
	22	amino acid sequence corresponding to SEQ ID NO:21
20	23	nucleotide sequence of pGEX and mcg7 junction
	. 24	amino acid sequence corresponding to SEQ ID NO:23
	25-36	mcg7-specific oligonucleotide
	37-45	mcg18-specific oligonucleotide

²⁵ Single and three letter abbreviations for amino acid residues are shown in Table 2.

TABLE 2

Amino Acid	Three-letter Abbreviation	One-letter Symbol
Alanine	Ala	Α ·
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
O Cysteine	Cys	С
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	Н
5 Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
) Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Тгр	W
Tyrosine	Tyr	Y
5 Valine	Val	V .
Any residue	Xaa	X

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a representation of the nucleotide sequence [SEQ ID NO:2] and corresponding amino acid sequence [SEQ ID NO:3] of mcg4.

5

Figure 2 is a representation of the alignment of the human MCG4 amino acid sequence with a translation of a partial murine expressed sequence tag (EST).

Figure 3 is a representation of the alignment of the human MCG4 amino acid sequence with a 10 translation of a partial nematode EST.

Figure 4 is a diagrammatic representation showing a predicted structure of MCG4 where H and C represent histidine and cysteine residues, respectively and X refers to any amino acid residue. Zn represent zinc atoms.

15

Figure 5 is a representation of sensitive sequence homology search of related cysteine-containing motifs in another *Caenorhabditis elegans* protein.

Figure 6 is a representation showing that a related cysteine containing motif is present in the 20 GATA-binding transcription factor from Saccharomyces pombe.

Figure 7 is a Northern blot showing expression of mcg4 in various cultured human cancer cell lines. Lanes 1-5, respectively, represent the hybridization signal from 15μg total RNA derived from various human cancer cell lines. Lanes 1-5, respectively, contain RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

Figure 8 is a representation of a partial alignment of mcg4 with human ESTs AA074703 and AA134788.

30

Figure 9 is a representation of the partial nucleotide sequence alignment between a human

(W32939) and mouse (AA242159) mcg4-like EST in the putative 5' UTR of the mcg4 cDNA. The putative initiation codon is underlined and the region upstream represents 5' UTR.

Figure 10 is a representation showing MacVector alignment of MCG4 with forward translations of ESTs AA134788 and AA074703. The nucleotide sequences are shown in Figure 8.

Figure 11 is a diagrammatic representation of the domains of MCG4

zinc finger consensus: CX₂HX₄CX₂CX₄HX₂CX₁₇CX₂CX₁₈HX₂CX₁₈CX₂C

acidic domain consensus: 9/34 amino acids negatively charged, 0/34 positively charged

basic domain consensus: 13/55 amino acids positively charged, 0/55 negatively charged

leucine zipper domain consensus: LX₆LX₆RX₆LX₆L

alternate "novel" leucine zipper-like motif where leucine would not be aligned along the one surface of an alpha helix domain: (aa261) LX₆LXLX₆LXLX₆L (aa 286).

15 **Figure 12** is a representation showing similarity of MCG7 with GEFs of various organisms.

Figure 13(a) is a representation of the nucleotide sequence [SEQ ID NO:4] and corresponding amino acid sequence [SEQ ID NO:5] of mcg7. Nucleotides 183-288 are an alternative spliced exon (shown in lower case).

20

Figure 13(b) is a representation of the partial nucleotide sequence [SEQ ID NO:6] and corresponding amino acid sequence [SEQ ID NO:7] of mcg7 but without the exon shown in Fig. 13(a). Amino acids have been numbered from the first methionine codon (underlined). The cDNA molecules of Fig. 13(a) and Fig. 13(b) differ by the inclusion and exclusion of the exon 25 of nucleotides 183-288.

Figure 14 is a representation showing a comparison between MCG7 and a homologue from *Caenorhabditis elegans* using the BESTFIT algorithm. In the figure, the following sequences are underlined:

30

la nematode DVDEEDEVEDIEF [SEQ ID NO:10]
lb human DVDGDGHISQEEF [SEQ ID NO:11]
nematode DHDRDGFISQEEF [SEQ ID NO:12]

1c human DQNQDGCISREEM [SEQ ID NO:13]

5 nematode DVDMDGQISKDEL [SEQ ID NO:14]

GUANINE NT BINDING REGION = BLOCKS DATABASE NO. BL00720B

2 human HFVHVAEKLLQLQNFNTLMAVVGGLSHSSISRLKETH[SEQID NO:15]

nematode KFVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTY

10 [SEQ ID NO:16]

DaG-PE BINDING DOMAIN = PROSITE DATABASE NO. PD0C00379

3 human HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC [SEQ ID NO:17]

15 nematode HNFHETTFLTPTTCNHCNKLLWGILRQGFKCKDCGLAVHSCCKSNAVAEC [SEQ ID NO:18]

Figure 15 is a representation of an alignment of human and a partial (5' UTR and partial coding sequence) murine mcg7 cDNA (GenBank Acc. No. W71787 and AA237373). The putative 20 initiation codon is underlined. The murine sequence represents a composite of 2 partial cDNA sequences from the EST database (accession numbers W71787 and AA237373). Nucleotide differences between human and murine sequences are shown in lower case lettering and identical residues are indicated with asterisks.

- 25 Figure 16 is a representation of further 5' nucleotide and corresponding amino acid sequence for human mcg7. Nucleotide positions 1-321 were derived from GenBank Acc. No. AC000134 and nucleotides 322 onwards from Fig. 13(a). Two in-frame initiation codons are underlined. Asterisks denote in-frame stop codons.
- 30 Figure 17 is a graphical representation of a GDP release assay. □ Experiment #1 (mean of duplicates). ♦ Experiment #2 (mean of duplicates). The exchange reaction contained 36pmols

of GST-MCG (N-terminally truncated; encoded by Construct B in Fig. 18) and 1.6-12.8 pmols of recombinant GST-N-Ras.GDP. Reaction time 6 mins.

Estimated reaction constants:

 $K_m = 2.1 \mu M$, $V_{max} = 37 p Mol/6 min/36 p Mol [Expt#1]$

5 $K_m = 1.5 \mu M$, $V_{max} = 30.3 p Mol/6 min/36 p Mol [Expt#2]$

Figure 18 depicts various recombinant plasmids containing partial or full-length mcg7.

Figure 19 is a representation of the nucleotide sequence [SEQ ID NO:8] and corresponding 10 amino acid sequence [SEQ ID NO:9] of mcg18.

Figure 20 is a representation showing that MCG18 has partial homology to E. coli DnaJ.

Figure 21 is a representation showing that MCG18 has homology to two *Caenorhabitis elegans* proteins.

Figure 22 is a representation showing that MCG18 has homology to a Saccharomyces pombe protein.

20 Figure 23 is a representation showing homology of MCG18 to a Drosophila virilis protein.

Figure 24 is a representation showing homology of MCG18 to human DnaJ proteins HDJ-2/HSDJ, HDJ-1/HSP40 and HSJ1.

25 Figure 25 is a representation of the nucleotide and corresponding amino acid sequence of murine *mcg18*.

Figure 26 is a representation of homology between human and murine MCG18.

30 Figure 27 depicts nucleotide sequences corresponding to the 5' untranslated region of human mcg18.

PCT/AU98/00380

Figure 28 depicts a Northern blot showing expression of mcg18 transcripts in total RNA isolated from various human cancer cell lines grown in culture. Lanes 1-5 respectively contain $15\mu g$ RNA from H69 lung carcinoma cells, JAM ovary carcinoma cells, BT20 breast carcinoma cells, HaCat transformed keratinocytes, T24 bladder carcinoma cells.

5

25

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having 5 homology to a regulator of gene expression or a derivative of said gene regulator.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding a regulator of gene expression wherein said regulator comprises a zinc finger domain of an (HC₃)₂ type.

Still more particularly, the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- 15 (i) a nucleotide sequence set forth in SEQ ID NO:2;
 - (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
 - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

The present invention also provides an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a guanine nucleotide exchange factor (GEF) or a derivative thereof.

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- 30 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or 7;

- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

5

Another aspect of the present invention contemplates an isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a heat shock protein or a heat shock-binding protein or a derivative thereof.

10

More particularly, the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides or a complementary form thereof selected from:

- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- 15 (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
 - (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
 - (iv) a nucleotide sequence capable of hybridizing under low stringency conditions at 42°C to the nucleotide sequence set forth in (i), (ii) or (iii).

20

Preferably, the percentage similarity is at least about 50%. More preferably, the percentage similarity is at least about 60%.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for hybridisation, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least

about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

The term "similarity" as used herein includes exact identity between compared sequences at the 5 nucleotide or amino acid level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which result in different amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational levels. Where there is non-identity at the amino acid level, "similarity" includes amino acids that are nevertheless related to each other at the structural, functional, biochemical and/or conformational 10 levels.

The present invention extends to nucleic acid molecules with percentage similarities of approximately 65%, 70%, 75%, 80%, 85%, 90% or 95% or above or a percentage in between.

15 The nucleic acid molecule of the present invention defined by SEQ ID NO:2 is hereinafter referred to as constituting the "mcg4" gene. The protein encoded by mcg4 is referred to herein as "MCG4" and has an amino acid sequence set forth in SEQ ID NO:3. The mcg4 gene is proposed to encode, in accordance with the present invention, a regulator of gene expression and comprises a novel zinc finger domain, (HC₃)₂. A regulator of gene expression includes a 20 transcription factor. Regulation may be at the level of nucleic acid:protein or protein:protein interaction.

The nucleic acid molecule of the present invention defined by SEQ ID NO:4 or 6 is hereinafter referred to as constituting the "mcg7" gene. The protein encoded by mcg7 is referred to herein 25 as "MCG7" and has an amino acid sequence set forth in SEQ ID NO:5 or 7 and is involved in signal transduction. The difference in the nucleotide and amino acid sequence is due to the presence or absence of an exon at nucleotides 183-288.

The nucleic acid molecule of the present invention defined by SEQ ID NO:8 is hereinafter 30 referred to as constituting the "mcg18" gene. The protein encoded by mcg18 is referred to herein as "MCG18" and comprises the amino acid set forth in SEQ ID NO:9.

The present invention extends to the naturally occurring genomic mcg4, mcg7 and mcg18 nucleotide sequences or corresponding cDNA sequences or to derivatives thereof. Derivatives contemplated in the present invention include fragments, parts, portions, mutants, homologues and analogues of MCG4, MCG7 or MCG8 or the corresponding genetic sequences. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG4, MCG7 or MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to mcg4, mcg7 or mcg18. "Additions" to the amino acid or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG4" or "mcg4", "MCG7" or "mcg7" or "MCG8" or mcg18" includes reference to all derivatives thereof including functional derivatives and immunologically interactive derivatives of MCG4, MCG7 or MCG18.

The mcg4, mcg7 and mcg18 of the present invention are particularly exemplified herein from humans and in particular from human chromosome 11q13.

15

The present invention extends, however, to a range of homologues from, for example, primates, livestock animals (eg. sheep, cows, horses, donkeys, pigs), companion animals (eg. dogs, cats) laboratory test animals (eg. rabbits, mice, rats, guinea pigs), reptiles, birds (eg. chickens, ducks, geese, parrots), insects, nematodes, eukaryotic microorganisms and captive wild animals (eg. deer, foxes, kangaroos). Reference herein to mcg4 and mcg18 or their respective proteins MCG4, MCG7 and MCG18 includes reference to these molecules of human origin as well as novel forms of non-human origin.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or

both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli*, *Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

- 5 Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg4 gene portion, which mcg4 gene portion is capable of encoding an MCG4 polypeptide or a functional or immunologically interactive derivative thereof.
- 10 Preferably, the mcg4 gene portion of the genetic construct is operably linked to a promoter in the vector such that said promoter is capable of directing expression of said mcg4 gene portion in an appropriate cell.

In addition, the *mcg4* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-Stransferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

20

It is proposed in accordance with the present invention that MCG4 is a transcription factor involved in gene regulation. Mutations in mcg4 may result in aberrations in gene regulation leading to the development of or a propensity to develop various types of cancer. In this regard, although not wishing to limit the present invention to any one hypothesis or mode of action, it 25 is proposed that mcg4 or its expression product may be involved in the tissue-specific or temporal regulation of particular genes.

A deletion or aberration in the mcg4 gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may

be determined by assaying for aberrations in the parents and/or proband of a subject under investigation.

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg4 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

10

Another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human mcg7 gene portion, which mcg7 gene portion is capable of encoding an mcg7 polypeptide or a functional or immunologically interactive derivative thereof.

15

Preferably, the mcg7 gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said mcg7 gene portion in an appropriate cell.

20 In addition, the mcg7 gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-Stransferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG7 is a GEF involved in signal transduction. Mutations in mcg7 or MCG7 may result in defective control of cell proliferation leading to the development of or a propensity to develop various types of cancer.

30

A deletion or aberration in the mcg7 gene may also be important in the detection of cancer or

a propensity to develop cancer. An aberration may be a homozygous mutation or a heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents of a subject under investigation.

5

According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg7 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.

Yet another aspect of the present invention contemplates a genetic construct comprising a vector portion and an animal, more particularly a mammalian and even more particularly a human 15 mcg18 gene portion, which mcg18 gene portion is capable of encoding an MCG18 polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the *mcg18* gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said *mcg18* gene portion 20 in an appropriate cell.

In addition, the *mcg18* gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

25

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

It is proposed in accordance with the present invention that MCG18 is a transcription factor 30 involved in protein folding, protein complex assembly and transit through subcellular compartments. MCG18 may also have a role in turnour suppression. Thus mutations in mcg18

may result in the development of or a propensity to develop various types of cancer.

A deletion or aberration in the *mcg18* gene may also be important in the detection of cancer or a propensity to develop cancer. An aberration may be a homozygous mutation or a 5 heterozygous mutation. The detection may occur at the foetal or post-natal level. Detection may also be at the germline or somatic cell level. Furthermore, a risk of developing cancer may be determined by assaying for aberrations in the parents and/or proband of the subject under investigation.

10 According to this aspect of the present invention, there is contemplated a method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg18 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or 15 a propensity to develop said condition.

The nucleotide substitutions, additions or deletions may be detected by any convenient means including nucleotide sequencing, restriction fragment length polymorphism (RFLP), polymerase chain reaction (PCR), oligonucleotide hybridization and single stranded conformation polymorphism analysis (SSCP) amongst many others. An aberration includes modification to existing nucleotides such as to modify glycosylation signal amongst other effects.

In an alternative method, aberrations in the mcg4, mcg7 and mcg18 genes are detected by screening for mutations in MCG4, MCG7 and MCG18, respectively.

25

A mutation in MCG4, MCG7 or MCG18 may be a single or multiple amino acid substitution, addition and/or deletion. The mutation in mcg4, mcg7 or mcg18 may also result in either no translation product being produced or a product in truncated form. A mutant may also be an altered glycosylation pattern or the introduction of side chain modifications to amino acid residues.

According to this aspect of the present invention, there is provided a method of detecting a condition caused or facilitated by an aberration in mcg4, mcg7 or mcg18 said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4, MCG7 or MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.

A particularly convenient means of detecting a mutation in MCG4, MCG7 or MCG18 is by use of antibodies.

- 10 Accordingly another aspect of the present invention is directed to antibodies to MCG4, MCG7 or MCG18 and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to MCG4, MCG7 or MCG18 or may be specifically raised to MCG4, MCG7 or MCG18 or derivatives thereof. In the case of the latter, MCG4, MCG7 or MCG18 or their derivatives may first need to be associated with a carrier molecule.
- 15 The antibodies to MCG4, MCG7 or MCG18 of the present invention are particularly useful as diagnostic agents.

For example, antibodies to MCG4, MCG7 or MCG18 and their derivatives can be used to screen for wild-type MCG4, MCG7 or MCG18 or for mutated MCG4, MCG7 or MCG18 molecules.

20 The latter may occur, for example, during or prior to certain cancer development. A differential binding assay is also particularly useful. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of normal MCG4, MCG7 or MCG18 levels or the presence of wild-type MCG4, MCG7 or MCG18 may be important for diagnosis of certain cancers or a predisposition for development of cancers or for monitoring 25 certain therapeutic protocols.

As stated above antibodies to MCG4, MCG7 or MCG18 of the present invention may be monoclonal or polyclonal or may be fragments of antibodies such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A "synthetic antibody" is considered herein to include fragments and hybrids of antibodies.

For example, specific antibodies can be used to screen for wild-type MCG4, MCG7 or MCG18 molecule or specific mutant molecules such as molecules having a certain deletion. This would be important, for example, as a means for screening for levels of MCG4, MCG7 or MCG18 in a cell extract or other biological fluid or purifying MCG4, MCG7 or MCG18 made by recombinant means from culture supernatant fluid or purified from a cell extract. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of wild-type MCG4, MCG7 or MCG18 or to a specific mutant phenotype or to a deleted or otherwise altered region.

15

Both polyclonal and monoclonal antibodies are obtainable by immunization of a suitable animal or bird with MCG4, MCG7 or MCG18 or its derivatives and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal or bird with an effective amount of MCG4, MCG7 or MCG18 or antigenic parts thereof or derivatives thereof, collecting serum from the animal or bird, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

25

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates a method for detecting MCG4, MCG7 or MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4, MCG7 or MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4, MCG7 or MCG18 complex to form, and then detecting said complex.

Preferably, the biological sample is a cell extract from a human or other animal or a bird.

The presence of MCG4, MCG7 or MCG18 may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4, 424,279 and 4,018,653. These include both single-site and two-site or "sandwich" assays of the non-competitive types, as well as traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

15

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into 20 contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigenlabelled antibody. Any unreacted material is washed away, and the presence of the antigen is 25 determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the 30 art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain MCG4, MCG7 or MCG18 including cell extract

or tissue biopsy. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

In the typical forward sandwich assay, a first antibody having specificity for the MCG4, MCG7 or MCG18 or an antigenic part thereof or a derivative thereof or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from room temperature to 37°C) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

20 An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-30 bound antibody. Detection may be either qualitative or quantitative. The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

WO 98/53061

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled 5 artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, betagalactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a 10 fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated. 15 usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

30

MCG7 or MCG18 or functional derivatives thereof. Such genetic constructs are also contemplated to be useful in modulating expression of specific genes in which mcg4, mcg7 or mcg18 is involved in tissue-specific or temporal regulation.

- 5 Accordingly, another aspect of the present invention is directed to a genetic construct comprising a nucleotide sequence encoding a peptide, polypeptide or protein and mcg4, mcg7 or mcg18 or a functional derivative or homologue thereof capable of modulating the expression of said nucleotide sequence.
- 10 As stated above, MCG18 is proposed to have a role in tumour suppression. Accordingly, it is further proposed in accordance with the present invention to use recombinant MCG18 in pharmaceutical preparations for treating arresting or otherwise ameliorating the effects of certain cancers.
- 15 Accordingly, another aspect of the present invention contemplates a method for treating, arresting or otherwise ameliorating the effects of a cancer in an animal or bird, said method comprising administering to said animal or bird an effective amount of MCG18 or a functional derivative thereof for a time and under conditions sufficient to treat, arrest or otherwise ameliorate the effects of said cancer.

20

The present invention, therefore, contemplates a pharmaceutical composition comprising MCG18 or a derivative thereof or a modulator of mcg18 expression or MCG18 activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to hereinafter as the "active ingredients". The active ingredients may also include anti-cancer agents or agents which facilitate actions of MCG18.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) and sterile powders for the extemporaneous preparation of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier may be a solvent medium containing, for example, water, ethanol, polyol (for example,

glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as licithin and by the use of superfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimersal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

10 Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with 20 the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 25 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 µg and 2000 mg of active compound.

30 The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter. A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium

phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release preparations and formulations.

The present invention also extends to forms suitable for topical application such as creams, lotions and gels.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired theraperatic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the

treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

The principal active ingredient is compounded for convenient and effective administration in 5 effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 µg to about 2000 mg. Expressed in proportions, the active compound is generally present in from about 0.5 µg to about 2000 mg/ml of carrier. In the case of compositions containing supplementary active ingredients, the dosages are 10 determined by reference to the usual dose and manner of administration of the said ingredients.

Effective amounts contemplated by the present invention include those amounts effective to ameliorate a condition. For example, it is envisaged that effective amounts would range from about 0.001 μg/kg body weight to about 100 mg/kg body weight. Alternatively, effective amounts of about 0.01 μg/kg body weight to about 10 mg/kg body weight or even 0.1 μg/kg body weight to about 1 mg/kg body weight. Administration may be per minute, hour, day, week, month or year or may only be a once off administration.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating mcg18 expression or MCG18 activity. The vector may, for example, be a viral vector.

As stated above, the present invention further contemplates a range of derivatives of MCG18. Derivatives include fragments, parts, portions, mutants, homologues and analogues of the MCG18 polypeptide and corresponding genetic sequence. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to MCG18 or single or multiple nucleotide substitutions, deletions and/or additions to the genetic sequence encoding MCG18. "Additions" to amino acid sequences or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to "MCG18" includes reference to all derivatives thereof including functional derivatives or MCG18 immunologically interactive derivatives.

Analogues of MCG18 contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

5

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with NaBH₄; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with NaBH₄.

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

20 Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide of sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids, contemplated herein is shown in Table 3.

TABLE 3

_	Non-conventional amino acid	Code	Non-conventional amino acid	Code
5	α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
	α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
	aminocyclopropane-	Cpro	L-N-methylasparagine	Nmasn
	carboxylate		L-N-methylaspartic acid	Nmasp
10	aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
	aminonorbornyl-	Norb	L-N-methylglutamine	Nmgln
	carboxylate		L-N-methylglutamic acid	Nmglu
•	cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
	cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
15	D-alanine	Dal	L-N-methylleucine	Nmleu
	D-arginine	Darg	L-N-methyllysine	Nmlys
	D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
	D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
	D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
20	D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
	D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
	D-isoleucine	Dile	L-N-methylproline	Nmpro
	D-leucine	Dleu	L-N-methylserine	Nmser
	D-lysine	Dlys	L-N-methylthreonine	Nmthr
25	D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
	D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
	D-phenylalanine	Dphe	L-N-methylvaline	Nmval
	D-proline	Dpro	L-N-methylethylglycine	Nmetg
	D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
30	D-threonine	Dthr	L-norleucine	Nle
	D-tryptophan	Dtrp	L-norvaline	Nva

	2	•	
-	•	7	-

	D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
	D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
	D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
	D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen
5	D-α-methylasparagine	Dmasn	α -methyl- α -napthylalanine	Manap
	D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
	D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
•	D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
	D-α-methylleucine	Dmleu	α-napthylalanine	Anap
	D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
	D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D-α-methylomithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	$D-\alpha$ -methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser
30	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp

	D-N-methyllysine	Dnmlys	N -methyl- γ -aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dommet
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
	N-methylglycine	Nala	D-N-methylphenylalanine	Drimphe
5	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
10	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	γ-aminobutyric acid	Gabu	N-(p-hydroxyphenyl)glycine	Nhtyr
	L-t-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L-α-methylalanine	Mala
15	L-α-methylarginine	Marg	L-α-methylasparagine	Masn
	L-α-methylaspartate	Masp	L-α-methyl-t-butylglycine	Mtbug
	L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
	L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
	L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
20	L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
	L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
	L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
	L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
25	L-α-methylserine	Mser	L-α-methylthreonine	Mthr
	L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr

WO 98/53061 PCT/AU98/00380

- 37 -

L-α-methylvaline Mval L-N-methylhomophenylalanine Nmhphe
N-(N-(2,2-diphenylethyl) Nnbhm N-(N-(3,3-diphenylpropyl) Nnbhe
carbamylmethyl)glycine carbamylmethyl)glycine
1-carboxy-1-(2,2-diphenyl- Nmbc
5 ethylamino)cyclopropane

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n=1 to n=6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N_α methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and 15 the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

Such analogues also apply in respect of MCG4 and MCG7.

20

The present invention further contemplates chemical analogues of MCG18 capable of acting as antagonists or agonists of MCG18 or which can act as functional analogues of MCG18. Chemical analogues may not necessarily be derived from MCG18 but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to mimic certain physiochemical properties of MCG18. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening.

The identification of MCG18 permits the generation of a range of therapeutic molecules capable of modulating expression of MCG18 or modulating the activity of MCG18. Modulators contemplated by the present invention includes agonists and antagonists of MCG18 expression. Antagonists of MCG18 expression include antisense molecules, ribozymes and co-suppression

molecules. Agonists include molecules which increase promoter ability or interfere with negative regulatory mechanisms. Agonists of MCG18 include molecules which overcome any negative regulatory mechanism. Antagonists of MCG18 include antibodies and inhibitor peptide fragments.

5

These types of modifications may be important to stabilise MCG18 if administered to an individual or for use as a diagnostic reagent.

Other derivatives contemplated by the present invention include a range of glycosylation variants from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

Another embodiment of the present invention contemplates a method for modulating expression of MCG18 in a human, said method comprising contacting the mcg18 gene encoding MCG18 with an effective amount of a modulator of mcg18 expression for a time and under conditions sufficient to up-regulate or down-regulate or otherwise modulate expression of mcg18. For example, a nucleic acid molecule encoding MCG18 or a derivative thereof may be introduced into a cell to facilitate protection of that cell from becoming cancerous.

20

Another aspect of the present invention contemplates a method of modulating activity of MCG18 in a human, said method comprising administering to said mammal a modulating effective amount of a molecule for a time and under conditions sufficient to increase or decrease MCG18 activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative of MCG18 or a chemical analogue or truncation mutant of MCG18.

The present invention is further described with reference to the following non-limiting Examples.

EXAMPLE 1

A human gene (designated mcg4) was identified on chromosome 11q13 that on the basis of sequence homology is predicted to encode a putative transcription factor of 310 amino acids 5 (Fig. 1). mcg4 is transcribed in several different cell lines (Fig. 7).

EXAMPLE 2

The expressed sequence tag (EST) database contains partial sequence data for the murine (Fig. 10 2) and nematode (Fig. 3) homologues of mcg4.

EXAMPLE 3

MCG4 contains a sequence of cysteine residues within the N-terminal region of the protein that resembles zinc-finger binding domains of a novel type, ie. (HC₃)₂ [Fig. 4].

EXAMPLE 4

Sensitive sequence homology searches reveal that related cysteine-containing motifs are present 20 in another *C. elegans* protein (Fig. 5) as well as the GATA-binding transcription factor from *S. pombe* (Fig. 6).

EXAMPLE 5

25 mcg4 will have commercial value due to its likelihood of encoding a novel transcription factor that is highly conserved amongst organisms, thus suggesting an integral role in gene regulation. mcg4 may also be involved in some way in tissue-specific or temporal regulation of certain genes, thus making it a potential target for modulating expression of those downstream effectors.

EXAMPLE 6

Nucleotide sequence data generated from cosmid clone cSRL-72c4 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) was aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul et al 1990) and was found to match numerous human and mouse entries (Table 4 and Figure 2). These matching ESTs were further used to identify overlapping entries in the EST database (Table 5). The nucleotide sequences of these human ESTs were complied using MacVector 4.2.1 software (IBI-Kodak) to produce the cDNA sequence shown in Figure 1. EST entries AA074703 and AA134788 are closely related at the nucleotide level to mcg4 and it is, therefore, likely that mcg4 is a member of a newly discovered gene family (Figure 8).

The cDNA sequence of mcg4 was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul et al, 1990) at 15 the National Center for Biotechnology Information (http://www.ncbi.nih.gov.nlm). As the protein appeared to be novel, a translation of the longest reading frame for the mcg4 cDNA was aligned to the EST database using the program TBLASTN, which performed a dynamic translation of the EST database in all 6 frames. The search results indicated that the nematode C. elegans had an MCG4-like protein (Figure 3), with the matching domains containing a spatial 20 sequence of Cysteine and Histidine residues which resembled a zinc-finger structure (Figure 4). The program BLASTP was used, therefore, to conduct sensitive searches of the protein databases for similar zinc-finger motifs. A weak match to the putative zinc-finger domain was observed for another protein from C. elegans (Figure 5) and a poorer match for the GATAbinding transcription factor from S. pombe (Figure 6). The putative initiation codon of human 25 mcg4 is not preceded by an in-frame stop codon and it is therefore possible that the cDNA described in Figure 1 is a truncated form. However, sequence alignment of human and mouse mcg4 ESTs showed a lower degree of nucleotide conservation prior to the assigned initiation codon, thus supporting the notion that the region represents the 5' UTR (Figure 9). To determine the expression pattern of mcg4, 15µg of the total cellular RNA (RNeasy Mini Kit, 30 Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% w/v MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer using 20 x SSC (Sambrook et al, 1989). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (32P-dCTP) cDNA probe (Church and Gilbert, 1984) for mcg4. After washes in 0.1 x SSC/0.1% w/v SDS at 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that mcg4 is expressed as a 1.6kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 7).

EXAMPLE 7

A human gene (designated *mcg7*) was identified and isolated from chromosome 11q13 which 10 encodes a protein that bears striking homology with guanine nucleotide exchange factors (GEFs) from a wide variety of organisms (Fig. 12).

EXAMPLE 8

15 The composite mcg7 cDNA sequence is at least 2.4kb in length and Figure 13(a) shows a predicted translation product of at least 609 amino acids beginning at methionine 120. An alternative start site due to alternate exon splicing (indicated in lower case) may yield a protein of 671 amino acids starting at methionine 58 (Fig. 13a).

20 EXAMPLE 9

An mcg7 homologue from C. elegans has been identified, the product of which is highly conserved with that of MCG7 (Fig. 14). There are several salient features of the protein which have been underlined in Fig. 14 - namely: a guanine nucleotide binding region, a diacylglycerol binding region, and "EF-hand"-calcium binding regions. In addition, there are several potential cAMP, protein kinase C, and casein kinase II phosphorylation sites, as well as a number of potential sites for glycosylation (not indicated).

EXAMPLE 10

30

A number of partial human and murine EST clones exist for mcg7. The GenBank database

5

contains a cDNA (Acc. no. Y12336) encoding a full-length open reading frame (ORF) for human mcg7 as well as a partial murine mcg7 ORF (Y12339). In addition, the complete genomic sequence of the human mcg7 gene is contained within GenBank entry AC000134.

EXAMPLE 11

The best characterised GEFs are members of the family of ras oncoproteins, which play a pivotal role in signal transduction and when mutated are responsible for tumour development. A variety of therapeutic regimes for cancer treatment have been designed to specifically interfere with the ras signalling pathways. There is potential, therefore that the product of mcg7 could also be a target for such clinical strategies.

EXAMPLE 12

The nucleotide sequence for mcg7 cDNA was extended 5' with genomic DNA sequence from Genbank accession number AC000134 (positions 1-321) and analysed for additional coding sequence 5' to the putative initiation codon (nt 681-683) (Fig. 16). An additional in-frame ATG occurs at position nt 495-497 when the alternatively splice exon (position nt 504-609) is present (also shown in Fig. 13(a)). This closely matches the Kozak consensus. When this exon is absent, then the ATG is not in-frame and other possible initiation codons are absent (resulting translation shown in lower case lettering) (also shown in Fig. 13(b)). Further evidence that the initiation codon at position nt 681-683 is the true initiation site is given in Figure 15.

Alignment of human and a partial murine mcg7 cDNA sequences is shown in Figure 15. The putative initiation codon is at position nt 360-362. Both murine ESTs appear to have an upstream in-frame stop codon at position nt 326-328, downstream of the differentially spliced exon and the sequence alignment thus suggests that this region represents the 5' UTR of mcg7.

Furthermore, similarity with the *C. elegans* homologue strongly suggest that the ATG codon at position at 360-362 encodes the N-terminus of MCG7.

EXAMPLE 13

Figure 17 shows data from experiments indicating that a truncated version of MCG7 when expressed as a GST fusion protein (construct B in Fig. 18) can function as a Ras-guanine nucleotide exchange factor. In brief, Ras (unprocessed and as a GST fusion protein) is loaded with ³H-GDP then incubated in the presence of excess cold GTP ± GST-MCG7. Full details of this assay can be found in Porfiri et al.

EXAMPLE 14

10

Nucleotide sequence data generated from cosmid clone cSRL-20h12 with the T7 primer (Promega, and Applied Biosystems Incorporated dye terminator sequencing kit) were aligned to the GenBank Expressed Sequence Tag (EST) database using the program BLASTN (Altschul et al, 1990) and was found to match GenBank entries T78563 (clone 113434) TO9103 (clone HIBBP12) and AA035643 (clone 471819). EST clones 113434 and 471819 were obtained from Genome Systems Inc. and these DNAs were sequenced on both strands with gene-specific primers (Table 5) to generate the cDNA sequence of mcg7 shown in Figures 13(a) and (b).

The cDNA sequence of *mcg*7 was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX (Altschul *et al*, 1990) and the coding region was assigned on the basis of showing homology to the *C. elegans* protein F25B3.3 (Figure 14). The *mcg*7 cDNA composite was suspected to contain a single nucleotide error that originated from clone 471819 and the correct nucleotide sequence was, therefore, sought by reverse transcription-polymerase chain reaction (RT-PCR) of the cDNA fragment from a human cDNA pool. Total RNA was extracted from a human lymphoblastoid cell line using an RNeasy Mini Kit (Qiagen). cDNA synthesis was conducted with the reverse transcriptase Superscript II RNaseH- (GIBCO, BRL) and random hexamers using the procedure recommended by the manufacturer (GIBCO, BRL). One fortieth of the cDNA mix was subjected to 35 cycles of PCR using the following cycling conditions: 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 90 seconds. The 50µl reaction mix consisted of 1x reaction buffer (Dade Scientific), 2mM dNTP mix, 20pmol of primers (see Table 6) MCG7UF (within the

variably spliced exon of Figure 13(b), between nucleotide positions 184-201) and SGCADRV2 (between nucleotide positions 866-846 of Figure 13(a)) and 10 units of Dynazyme (Dade Scientific). The resulting PCR product was cloned into the pGEM-T vector (Promega) using standard methodology and sequenced using gene-specific primers. The correct nucleotide sequence of mcg7 (as shown in Figure 13(a)) matches that of the recently release GenBank entry Y12336. A partial mouse mcg7 cDNA sequence can also be found in GenBank entry Y12339.

EXAMPLE 15

10 The coding sequence of mcg7 was cloned into vectors for expression in both bacterial and mammalian cells. In addition to the full-length constructs, the deletion constructs shown in Figure 18 were designed to retain the guanine nucleotide exchange (GEF) domain. For prokaryotic expression, the mcg7 coding region was inserted downstream of and in-frame with the Sj26 cassette of the pGEX (Pharmacia) series of vectors (Smith and Johnson, 1988) using standard cloning techniques (Sambrook et al, 1989). For mammalian expression, the mcg7 coding sequence was first myc-tagged at the N-terminus and then ligated into the expression vector pc Exv-n using standard cloning techniques. Ligation junctions of the constructs were sequences as the cloning strategies inadvertently changed or introduced additional amino acids as shown below.

20

Construct (A): EST clone 113434 was digested with *Apa*I (Figure 13(a), nucleotide positions 1022 to >2416 (within the vector)), blunt-ended with T4 DNA polymerase according to the specifications of the manufacturer (New England Biolab) and ligated into the *Sma*I site of pGEX-3X.

25

Sequence of the pGEX and mcg7 (underlined) junction:

pGEX-3X

mcg7 (1022)

Sj26 ... GGG ATC CCC CTG GTC [SEQ ID NO:19]

additional amino acids Gly Ile Pro

30

Construct (B): EST clone 113434 was digested with EcoRI (Figure 13(a), nucleotide

WO 98/53061 PCT/AU98/00380

positions <695 (within the vector) to 1711) and ligated into the *EcoRI* site of pGEX-1.

- 45 -

Sequence of the pGEX and mcg7 (underlined) junction:

pGEX-1

mcg7 (695)

5 Sj26 ... GAA TTC GGC ACG AGC CGA CGG [SEQ ID NO:20] additional amino acids Glu Phe Gly Thr Ser

Construct (C): full-length mcg7: The pGEM-T clone containing the 5' end of the mcg7 coding region was digested with ApaI (subsequently blunt-ended with T4 DNA polymerase) and BstXI 10 to liberate the fragment between nucleotide positions 336 and 830 of Figure 13(a). Clone 113434 was digested with BstXI and HindIII (vector derived) to liberate a fragment between nucleotide positions 830 > and 2416 (vector derived) of Figure 13(a). A pGEM-11zf vector (Promega) containing the myc-tag was digested with ApaI (subsequently blunt-ended with T4 DNA polymerase) and *Hind*III, and ligated with the 2 inserts described above.

15

Sequence of the myc-tag/mcg7 junction [SEQ ID NOs:21/22]:

-----myc-tag----vector BamHI mcg7 5' UTR (337) ATGGAGCAGAAGCTGATCTCCGAGGAGGACCTG CCCGGGGCAGCTggatccG CAGCCCACCCCGCGCCGCCGCCATC 20 MEQKLISEEDL PGAAGS AAHPAPAAM -----additional amino acids-----

The myc-tagged full-length mcg7 insert in pGEM-11zf was then excised with SacI and HindIII (both vector derived) and directionally cloned into the mammalian expression vector pEXV 25 (Beranger et al, 1994).

Construct (D): Construct (C) in pGEM-11zf was sequentially digested with *Hind*III (this site was subsequently blunt-ended with T4 DNA polymerase) then BamHI, and ligated into pGEX-2T digested with BamHI and SmaI. Digestion with BamHI, and ligated into pGEX-2T digested 30 with BamHI and SmaI. Digestion with BamHI removed the myc-tag of Construct (C).

Sequence of the pGEX and mcg7 [SEQ ID NO:23/24] (underlined) junction:

-----additional amino acids------

pGEX-2 BamHI mcg7 (337)
Sj26 ... gga tcc GCA GCC CAC CCC GCG CCG GCC GCC ATG
Gly Ser Ala Ala His Pro Ala Pro Ala Met

5

EXAMPLE 16

Overnight bacterial cultures containing the pGEX plasmid were used to inoculate 500ml of Luria Broth media containing $50\mu g/ml$ ampicillin. The cultures were grown to an OD of ~0.8 and then 10 induced with 1mM of IPTG for up to 3 hours at 37°C. The bacteria were pelleted and resuspended in 15 ml of STE buffer (10mM Tris pH 8.0, 150 mM NaCl and 1mM EDTA) with 1 mg/ml lysozyme. The mixture was left on ice for more than 1 hour and subsequent steps were performed at 4°C. Protease inhibitors aprotinin, pepstatin and leupeptin were added at final concentrations of 25µg/ml, prior to the addition of Triton-X-100 (2% v/v final) and n-lauroyl 15 sarcosine (1.5% w/v final). The lysate was sonicated for ~1 minute and pelleted at 14,000 x g for 15 minutes. 100 μ l of 50% w/v glutathione-sephadex bead slurry (in PBS) was added per ml of supernatant. Following a 30 minute incubation at 4°C, the beads were washed three times with NETN (20mM Tris-HC1 pH 8.0, 100mM NaCl, 1mM EDTA, 0.5% NP40), once with NETN-HS (equivalent to NETN but with 1M NaCl), and once in NETN. The bound protein 20 was directly analysed by SDS-polyacrylamide gel electrophoresis (PAGE) as described below or the bound protein was eluted from the beads with the following elution buffer (50mM Tris pH 8.0, 150mM NaCl, 5mM MgCl₂, 1mM DTT, 10mM reduced glutathione) for use in GDP release assays.

25

EXAMPLE 17

Twenty microlitres of GST-sepharose-bound MCG7 were added to an equal volume of 2 x 30 sample loading dye (100mM Tris pH6.8, 2% v/v mercaptoethanol, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol), boiled for 5 min and loaded onto a 7.5% w/v SDS-PAGE gel (Sambrook *et al*, 1989). The Coomassie brilliant blue stained gel (Sambrook *et al*, 1989)

typically displayed a protein doublet, running between 87-95 kDa consisting of the MCG7-GST fusion and a slightly smaller, co-purified contaminating E. coli protein of ~105kDa. The calculated molecular weight of full-length MCG7 is 77.5 kDa (Construct (D)) and the GST component has a molecular weight of 26kDa, hence, the recombinant protein runs slightly smaller than predicted. A Western blot of the same gel probed with anti-GST antibody yields an MCG7-specific band at the same position as that of the stained gel.

EXAMPLE 18

10 Assumptions: (a) GST-Ras molecular weight = 50 kD; (b) Concentration of GST-Ras solution = 1mg/ml = 20μM; (c) [³H]-GDP is 1mCi/ml and 13.3Ci/mmol, therefore [H]-GDP concentration = 75 μM and 1pmol [³H]-GDP=15,466 cpm; (d) Elution buffer = Buffer E = 20 mM Tris-Cl, pH7.5; 50mM NaCl; 5mM MgCl₂; 1mM DTT (added just before use). Buffer E + BSA= Buffer E+1mg/ml BSA (added just before use).

15

Mix together, in the following order and mix well after each addition:

10μl (=10μg) GST-Ras (@1mg/ml in Buffer E), 463μl Buffer E + BSA, 7μl [³H]-GDP, 10ml

490 μM EDTA. Incubate @ RT for 10 min. Add 10μl 0.5 M MgCl₂ and mix well. Incubate

@ RT for 10 min. Place on ice. During the first incubation the excess EDTA concentration is

20 5mM, during the second incubation the excess Mg concentration is 5mM. The [³H]-GDP concentration is 1μM and the final concentration of GST-Ras is 400nM. Thus 20ml of the final mix will contain 8pmol of GST-Ras protein. Specific activity of GDP is 15,446 cpm/pmol x

(1/1.4) = 11,047 cpm/pmol.

25

EXAMPLE 19

Exchange Ras with labelled GDP as above. Add unlabelled GTP (stock = 100mM, pH7) to 1 mM. Adjust Mg concentration by adding 5μl 0.5 EDTA to labelled Ras, 5μl 0.5M EDTA to 500μl MCG7, and 5μl 0.5M EDTA to 500μl Buffer E + BSA. On ice set up microfuge tubes with 40μl Ras-GDP (in triplicate) with 40μl MCG7 or Buffer E + BSA (control). Transfer tubes to heat block @ 25°C and incubate for 10, 20 or 30 min. Stop exchange reactions with 1ml of

ice cold buffer E and place on ice. Pre-soak nitrocellulose filters, pore size 45µm, in Buffer E. Assemble the vacuum manifold apparatus (Millipore) with wet filters and plug the wells with rubber bunds. Switch on the vacuum pump. Remove the first plug, aliquot the sample and once it has been sucked through, wash the filter with 10ml of ice cold Buffer E. Remove next plug etc and continue round the manifold. Take manifold apart. Pin the filters to a pin board reserved for [³H]. Air dry. Take up in 4ml scintillation fluid and count. These studies have been carried out with a truncated MCG7-GST fusion protein (amino acids 341 of Figure 13a to stop encoded within construct B).

10

EXAMPLE 20

A human gene was identified from chromosome 11q13 that encodes a new member of the DnaJ family of proteins (designated MCG18). This gene (mcg18) is expressed as an ~1.4kb mRNA (Fig. 28) and is predicted to encode a 241 amino acid product (Fig. 19).

15

EXAMPLE 21

MCG18 has partial homology to *E. coli* dnaJ and other human DnaJ family members in that it contains the J domain (Fig. 20).

20

EXAMPLE 22

MCG18 has greatest homology to functionally undefined proteins from *C. elegans* (Fig. 21) and *S. pombe* (Fig. 22) that also feature the J domain but maintain sequence similarity through the central and C-terminal regions of the proteins.

EXAMPLE 23

The J domain is proposed to mediate interaction with heat shock protein (Hsp70) 70 and consist 30 of some 70 amino acids, frequently located at the N-terminus of the protein. One of these proteins, tumorous imaginal discs (Tid58) from *Drosophila virilis* (Fig. 23) functions as a

tumour suppressor.

EXAMPLE 24

5 A comparison of homology between MCG18 and human DnaJ proteins HDJ-2/H5DJ, HDJ-1/HSP40 and HSJ1 is shown in Fig. 24.

EXAMPLE 25

- 10 During the sequence characterisation of the VRF/VEGFB promoter region on cosmid CLGW4 [Grimmond et al, 1996], which maps to chromosome 11q13 the inventors identified a sequence that exactly matched numerous human and mouse expressed sequence tags (ESTs) in the EST database from a gene which we designated mcg18. EST clones for human (GenBank accession number T69741, clone 108172; accession number H40901, clone 177008) and mouse mcg18
 15 (accession number W34884, clone 350966; accession number W64183, clone 385535) were obtained from Genome Systems Inc. and sequenced with the gene-specific primers shown in Table 7. The EST clones listed in Table 8 were also utilised in generating the full-length coding sequence for human (Figure 19) and mouse (Figure 25) mcg18. The EST database also contained mcg18 cDNA entries that were alternately (or partially) spliced, and in order to understand their ability to encode new polypeptides, the gene structure of mcg18 was determined by sequencing human and mouse genomic templates with gene-specific primers.
- Genomic fragments containing the human [Grimmond et al, 1996] and murine genes [Townson et al, 1996] have been previously reported. Cosmid CLGW4 contains the entire human gene and λ121 contains the entire mouse gene, as determined by direct sequencing of the templates with the oligonucleotides listed in Table 7. Plasmids containing sub-fragments of λ121 and cosmid CLGW4 were prepared using plasmid purification kits (Qiagen) and sequenced as described previously [Grimmond et al, 1996; Townson et al, 1996] using primers designed against cDNA and genomic sequences. The BLAST suite of programs [Altschul et al, 1990] was used to compare the sequence data against the nucleotide and protein databases at the National Center for Biotechnology Information (http://www.ncbi.nih.gov.nlm). The sequence

data were compiled using MacVector 4.2.1 software (IBI-Kodak). ClustalW sequence alignments [Thompson *et al*, 1994] were conducted using the Australian National Genome Information Service computer faculty at the University of Sydney, Australia.

- 5 The cDNA sequence of human mcg18 (Figure 19) was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program BLASTX [Altschul et al, 1990] and the coding region was identified on the basis of showing homology to the DnaJ family of proteins (Figure 20). The DnaJ domain is encoded within the longest open reading frame and the assigned initiation codon is preceded by an in-frame stop codon (Figure 27). Similar database search results were obtained for the mouse mcg18 cDNA, and the alignment of human and mouse protein sequences is shown in Figure 26. MCG18 has greatest homology to gene products from C. elegans (Figure 21) and S. pombe (Figure 22). Although it shares a similar J-domain, MCG18 does not contain other domains described for the tumour suppressor gene from D. virilis (Figure 23), nor is it a homologue of other reported human J-domain-containing proteins (Figure 24).
- To determine the expression pattern of mcg18, $15\mu g$ of total cellular RNA (RNeasy Mini Kit, Qiagen) from various human cell lines grown in culture were electrophoresed through 1.2% MOPS/formaldehyde gels and blotted onto nylon membranes (Amersham) by capillary transfer using 20 x SSC (Sambrook et al, 1986). Filters were subsequently UV-fixed and hybridised overnight at 65°C to a radiolabelled (^{32}P -dCTP) cDNA probe (Church and Gilbert, 1984) for mcg18. After washes in 0.1 x SSC/0.1% w/v SDS for 65°C for 1 hour, the filters were air-dried and exposed to X-ray film. This Northern analysis showed that mcg18 is expressed as a 1.4kb message in numerous tissues including breast, ovary, bladder, lung and keratinocytes (Figure 28).

- 51 -

TABLE 4

ESTs matching mcg4

accession number	seq. run	organism	score	E value	N
gb AA399110 AA39911	0 zt89e06.sl	Soares testis NHT Homo sa	1136	4.0e-168	2
gb N39612 N39612	yy51g06.s1	Homo sapiens cDNA clone 2	1521	5.3e-168	4
gb AA514406 AA51440	nf57d01.s1	NCI_CGAP_Co3 Homo sapiens	931	5.5e-166	3
gb AA544946 AA54494		Soares mouse mammary glan	1207	8.4e-164	2
gb AA450076 AA450076		Soares total fetus Nb2HF8	691	2.3e-160	4
gb AA535731 AA535731	nf88f07.sl	NCI_CGAP_Co3 Homo sapiens	796	3.5e-158	4
gb W79710 W79710		Soares fetal heart NoHH19	1644	1.le-157	4
gb AA503531 AA503531	ne47e08.sl	NCI_CGAP_Col Homo sapiens	736	4.0e-156	4
gb AA450132 AA450132		Soares total fetus Nb2HF8	1955	3.9e-155	i
gb AA398068 AA398068		Soares testis NHT Homo sa	1315	5.4e-148	2
gb W60405 W60405		Soares fetal heart NbHH19	1022	1.8e-139	4
gb W81382 W81382		Soares fetal heart NbHH19	605	3.5e-125	5
gb AA047617 AA047617		Soares fetal heart NbHH19	922	4.6e-125	2
gb AA282175 AA282175	zt02d03.s1	NCI_CGAP_GCB1 Homo sapien	1577	2.0e-123	ī
gb AA242159 AA242159	my30d04.rl	Barstead mouse pooled org	866	7.7e-117	2
gb AA068680 AA068680	mm61a05.rl	Stratagene mouse embryoni	1280	1.6e-98	1
gb W46766 W46766		Soares senescent fibrobla	506	9.6e-92	3
gb N93704 N93704		Soares fetal lung NbHL19W	584	9.0e-91	4
gb AA155210 AA155210	mr98e01.rl	Stratagene mouse embryoni	840	7.6e-87	2
ap yy366055 yy366055	EST76915 Pi	neal gland II Homo sapien	1077	2.4e-81	1
gb AA037691 AA037691	zk34h12.s1	Soares pregnant uterus Nb	949	2.1e-80	2
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor	1016	3.le-76	1
qpi C00e3e C00e3e	HUMGS000825	1, Human Gene Signature,	1009	1.2e-75	1
gb T98249 T98249	ye59a07.s1	Homo sapiens cDNA clone 1	998	6.7e-75	1
gb W21588 W21588	zb51c04.rl	Soares fetal lung NbHL19W	484	1.1e-69	4
gb H32171 H32171		attus sp. cDNA 5' end.	828	1.le-60	1
gb AA108092 AA108092	mm89e06.rl	Stratagene mouse embryoni	782	1.3e-60	2
gb AA017857 AA017857		Soares mouse placenta 4Nb	665	2.5e-60	2
gb AA037690 AA037690	zk34h12.rl	Soares pregnant uterus Nb	540	9.4e-53	2
gb AA531006 AA531006		NCI_CGAP_Pr22 Homo sapien	535	5.4e-48	2
gb N46760 N46760	yy51g06.rl 1	Homo sapiens cDNA clone 2	665	9.5e-47	1
gb W23584 W23584	zc71d03.s1	Soares fetal heart NbHH19	457	1.8e-44	2
gb W42214 W42214	mc69h09.rl	Soares mouse embryo NbME1		1.3e-38	3
gb AA244877 AA244877	mx25a04.rl :	Soares mouse NML Mus musc		2.9e-25	1
gb W32939 W32939	zc07h03.rl 5	Soares parathyroid tumor		4.8e-18	ī

- 52 -

TABLE 5
ESTs matching AA074703 (mcg4-related cDNA)

Database: Non-redundant Database of GenBank EST Division
1,222,625 sequences; 449,352,662 total letters.

Smallest

Sum

			High	Probabil:	ity
Sequences producing F	High-scoring	Segment Pairs:	Score	P (N)	N
accession number	seq. run	organism	score	E value	N
gb AA074703 AA074703	zm76g07.rl	Stratagene neuroepitheli	2071	4.0e-167	1
gb AA068680 AA068680	mm61a05.r1	Stratagene mouse embryon	1270	4.4e-145	4
gb AA134788 AA134788	zm81g02.rl	Stratagene neuroepitheli	946	1.3e-144	5
gb AA399110 AA399110	zt89e06.s1	Soares testis NHT Homo s	520	8.7e-119	6
gb N39612 N39612	yy51g06.s1	Homo sapiens cDNA clone	582	9.6e-110	7
gb AA282175 AA282175	zt02d03.sl	NCI_CGAP_GCB1 Homo sapie	771	9.4e-80	3
gb W81382 W81382	zd86f01.s1	Soares fetal heart NoHH1	329	1.6e-75	6
gb AA544946 AA544946	vk38e02.r1	Soares mouse mammary gla	644	9.6e-63	2
gb W35374 W35374	zc07h03.s1	Soares parathyroid tumor	294	4.5e-42	4
gb W57106 W57106	md57c12.r1	Soares mouse embryo NbME	394	1.9e-30	2
gb AA244877 AA244877	mx25a04.r1	Soares mouse NML Mus mus	162	2.1e-27	4
gb AA017857 AA017857	mh44d10.r1	Soares mouse placenta 4N	230	3.7e-23	3
gb AA531006 AA531006	nj07b11.s1	NCI_CGAP_Pr22 Homo sapie	139	2.3e-19	3
gb H32171 H32171	EST107015 R	attus sp. cDNA 5' end.	207	2.6e-10	2
gb W79710 W79710	zd86f01.r1	Soares fetal heart NbHHl	157	0.0073	1

TABLE 6

mcg7-specific oligonucleotides

5	name	sequence (5' to 3')	SEQ ID NOs.
	M1044R	GGA CAA AGT GTG TGA TGA ACC	SEQ ID NO:25
	MCG7-GEF-REV2	CTC ATC CTC CGT CTG ATA CTG	SEQ ID NO:26
	M7R	GTA GAT GTG GAT CAG CTT GG	SEQ ID NO:27
	MCG7 CA FOR	AGG TGG AGA ATG GTC AAGG	SEQ ID NO:28
10	MCG7-GEF-REV	GTC ATA GTC TGT CTC CTA CT	SEQ ID NO:29
	MCG7 GEF FOR	ACA TAG ACA GCG TGC CTA CC	SEQ ID NO:30
	MCG7-PKC-REV	TAC AAC CTT AGG GAC ACC AG	SEQ ID NO:31
	MCG7-PKC-FOR	TGC TGA GCC TGC TCA CGG TG	SEQ ID NO:32
	T09103F	CAA GTG AAC AGC ACG TCC	SEQ ID NO:33
15	M7F	GAC TAT CTC AAG GAC CAG CTG	SEQ ID NO:34
	MCG7UF	GGT TCG GTC CGA GCC CGG	SEQ ID NO:35
	SGCADRV2	GGA GCG ATA CTC CAA GTA GGT	SEQ ID NO:36

TABLE 7
mcg18-SPECIFIC OLIGONUCLEOTIDES

	name	sequence 5' to 3'
5	HVESTF	AGC GGG CCA GGC CCC TTC [SEQ ID NO:37]
	HV195F	CAT CCT GGT CCA ATG CGC TC [SEQ ID NO:38]
	HV387F2	GCA CTG AGG AAG TTA AAC GAG C [SEQ ID NO:39]
	HV408R	GCT CGT TTA ACT TCC TCA GTG C [SEQ ID NO:40]
	EXONIREV	GCT CAG CTC CAC AAA GCG GCT [SEQ ID NO:41]
10	HVEST426F	ACC AGC TCC GCT CAG GTA G [SEQ ID NO:42]
	HVEST623R	TCC AGG AGC TGT GTG TTT GG [SEQ ID NO:43]
	SGVESTF3	CCA GTT TCA CAG CGT GAG G [SEQ ID NO:44]
	HVEST631R	CAG CAT GAG GAG GAG GCA G [SEQ ID NO:45]

TABLE 8
EST CLONE SEQUENCES USED TO GENERATE HUMAN AND MOUSE

mcg18 cDNA SEQUENCE COMPOSITES

EST clone number	organism	GenBank accession number
1g2815	human	D45683
001-T2-18	human	F17225
273748	human	N37043
177008	human	H40901 and H40939
258011	human	N30776
276887	human	N44004
108172	human	T69741
307529	human	W21083 and W32579
342027	human	W60283
354288	mouse	W44038
350966	mouse	W348844
426261	mouse	AA002868
368185	mouse	W53911
385535	mouse	W64183
404472	mouse	W82959
406437	mouse	W83482

BIBLIOGRAPHY

- 1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) J. Mol. Biol. 215: 403-410.
- 2. Church, G., and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 18: 1991-1995.
- 3. Porfiri et al. J. Biol. Chem. 269: 22672-22677 (1994).
- 4. Sambrook, J., Frtisch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbour Laboratory, Cold Spring Harbour, NY, USA.
- 5. Smith, D.B., and Johnson, K.S. (1988) Gene 67: 31-40.
- 6. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) *Nucleic Acids Res.* 22: 4673-4680.
- 7. Beranger, F., Paterson, H., Powers, S., de Gunzburg, J. and Hancock, J.F. (1994)

 Molecular and Cellular Biology 14: 744-758.
- 8. Grimmond, S., Lagercrantz, J., Drinkwater, C., Silins, G., Townson, S., Pollock, P., Gotley, D., Carson, E., Rakar, S., Nordenskjöld, M., Ward, L., Hayward, N., and Weber, G (1996) Genome Res. 6: 124-131.
- 9. Townson, S., Lagercrantz, J., Grimmond, S., Silins, G., Nordenskjöld, Weber, G., and Hayward, N. (1996) *Biochem. Biophys. Res. Commun.* 220: 922-928.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: (OTHER THAN US):

The Council of The Queensland Institute of

Medical Research

(US ONLY): HAYWARD Nicholas, SILINS Ginters, GRIMMOND Sean, GARTSIDE Michael and HANCOCK, John

- (ii) TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
- (iii) NUMBER OF SEQUENCES: 45
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: DAVIES COLLISON CAVE
 - (B) STREET: 1 LITTLE COLLINS STREET
 - (C) CITY: MELBOURNE
 - (D) STATE: VICTORIA
 - (E) COUNTRY: AUSTRALIA
 - (F) ZIP: 3000
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT INTERNATIONAL
 - (B) FILING DATE: 22-MAY-1998
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PO6973
 - (B) FILING DATE: 23-MAY-1997
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PO6974
 - (B) FILING DATE: 23-MAY-1997
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PO6972
 - (B) FILING DATE: 23-MAY-1997

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1459
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1460
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: PP1458
- (B) FILING DATE: 22-JAN-1998
- (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: HUGHES, DR E JOHN L
- (C) REFERENCE/DOCKET NUMBER: EJH/AF

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: +61 3 9254 2777
- (B) TELEFAX: +61 3 9254 2770
- (C) TELEX: AA 31787

(2)	INFORMATION	FOR	SEO	ID	NO:	1	:
-----	-------------	-----	-----	----	-----	---	---

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Cys Xaa Xaa Cys Xaa Gly Xaa Gly

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1242 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single

 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 30..959
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

			-					_								
TCA	GTAA	ACA (CAGA	GACT	GG G(GATC	GATC					AAG Lys 5				53
	AAG Lys 10															101
	CAC His															149
	CAA Gln															197
	ATA Ile															245
CTC Leu	TTT Phe	CAC His 75	TGG Trp	GCC Ala	TGC Cys	CTC Leu	AAT Asn 80	GAA Glu	CGT Arg	GCT Ala	GCC Ala	CAG Gln 85	CTA Leu	CCC Pro	CGA Arg	293
AAC Asn	ACG Thr 90	GCA Ala	CCT Pro	GCC Ala	GGC Gly	TAT Tyr 95	CAG Gln	TGC Cys	CCC Pro	AGC Ser	TGC Cys 100	AAT Asn	GGC Gly	CCC Pro	ATC Ile	341
TTC Phe 105	CCC Pro	CCA Pro	ACC Thr	AAC Asn	CTG Leu 110	GCT Ala	GGC Gly	CCC Pro	GTG Val	GCC Ala 115	TCC Ser	GCA Ala	CTG Leu	AGA Arg	GAG Glu 120	389

AAG Lys	CTG Leu	GCC Ala	ACA Thr	GTC Val 125	AAC Asn	TGG Trp	GCC Ala	CGG Arg	GCA Ala 130	GGA Gly	CTG Leu	GGC Gly	CTC Leu	Pro 135	CTG Leu	437
ATC Ile	GAT Asp	GAG Glu	GTG Val 140	GTG Val	AGC Ser	CCA Pro	GAG Glu	CCC Pro 145	GAG Glu	CCC Pro	CTC Leu	AAC Asn	ACG Thr 150	TCT Ser	GAC Asp	485
TTC Phe	TCT Ser	GAC Asp 155	TGG Trp	TCT Ser	AGT Ser	TTT Phe	AAT Asn 160	GCC Ala	AGC Ser	AGT Ser	ACC Thr	CCT Pro 165	GGA Gly	CCA Pro	GAG Glu	533
														GCC Ala		581
														ATC Ile		629
														TAT Tyr 215		677
ACG Thr	CGG Arg	GAT Asp	GAT Asp 220	GAC Asp	CGG Arg	ACA Thr	CCA Pro	GGC Gly 225	CTC Leu	CAT His	GGA Gly	GAC Asp	TGT Cys 230	GAC Asp	GAT Asp	725
														CTG Leu		7 73
AGG Arg	AGC Ser 250	CGG Arg	GCT Ala	GGG Gly	TCT Ser	CGG Arg 255	AAG Lys	CGG Arg	CCG Pro	CTG Leu	ACC Thr 260	CTG Leu	CTC Leu	CAG Gln	CGG Arg	821
GCG Ala 265	GGG Gly	CTG Leu	CTG Leu	CTA Leu	CTC Leu 270	TTG Leu	GGA Gly	CTG Leu	CTG Leu	GGC Gly 275	TTC Phe	CTG Leu	GCC Ala	CTC Leu	CTT Leu 280	869
GCC Ala	CTC Leu	ATG Met	TCT Ser	CGC Arg 285	CTA Leu	GGC Gly	CGG Arg	GCC Ala	GCA Ala 290	GCT Ala	GAC Asp	AGC Ser	GAT Asp	CCC Pro 295	AAC Asn	917
								ATC Ile 305						TGA *		962
GCC	CCCT'	TGC	TTGT	GGCT.	AG G	CCAG	CCTA	G GA	TGTG	GGTT	CTG	TGGA	GGA (GAGG	CGGGGT	1022
AAT	GGGG	AGG	CTGA	GGC.	AC C	TCTT	CACT	G CC	CCTC	TCCC	TCA	AGCC'	TAA	GACA	CTAAGA	1082
CCC	CAGA	ccc .	AAAG	CCAA	GT C	CACC.	AGAG	r gg	CTCG	CAGG	CCA	GGCC'	TGG .	AGTC	CCCGTG	1142
GGT	CAAG	CAT	TTGT	CTTG.	AC T	TGCT	TTCT	c cc	GGGT	CTCC	AGC	CTCC	GAC	CCCT	CGCCCC	1202
ATG.	AAGG.	AGC	TGGC.	AGGT	GG A	AATA	AACA	A CA	ACTT	TATT						1242

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 310 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly Leu Cys Lys Cys Pro Lys Arg Lys Val Thr Asn Leu Phe Cys Phe Glu His Arg Val Asn Val Cys Glu His Cys Leu Val Ala Asn His Ala Lys Cys Ile Val Gln Ser Tyr Leu Gln Trp Leu Gln Asp Ser Asp Tyr Asn Pro Asn Cys Arg Leu Cys Asn Ile Pro Leu Ala Ser Arg Glu Thr Thr Arg Leu Val Cys Tyr Asp Leu Phe His Trp Ala Cys Leu Asn 65 70 75 80 Glu Arg Ala Ala Gln Leu Pro Arg Asn Thr Ala Pro Ala Gly Tyr Gln
85 90 95 Cys Pro Ser Cys Asn Gly Pro Ile Phe Pro Pro Thr Asn Leu Ala Gly 105 Pro Val Ala Ser Ala Leu Arg Glu Lys Leu Ala Thr Val Asn Trp Ala Arg Ala Gly Leu Gly Leu Pro Leu Ile Asp Glu Val Val Ser Pro Glu Pro Glu Pro Leu Asn Thr Ser Asp Phe Ser Asp Trp Ser Ser Phe Asn Ala Ser Ser Thr Pro Gly Pro Glu Glu Val Asp Ser Ala Ser Ala Ala Pro Ala Phe Tyr Ser Arg Ala Pro Arg Pro Pro Ala Ser Pro Gly Arg 185 Pro Glu Gln His Thr Val Ile His Met Gly Asn Pro Glu Pro Leu Thr His Ala Pro Arg Lys Val Tyr Asp Thr Arg Asp Asp Asp Arg Thr Pro Gly Leu His Gly Asp Cys Asp Asp Asp Lys Tyr Arg Arg Arg Pro Ala 225 230 235 Leu Gly Trp Leu Ala Arg Leu Leu Arg Ser Arg Ala Gly Ser Arg Lys Arg Pro Leu Thr Leu Leu Gln Arg Ala Gly Leu Leu Leu Leu Gly Leu Leu Gly Phe Leu Ala Leu Leu Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Ala Asp Ser Asp Pro Asn Leu Asp Pro Leu Met Asn Pro His Ile Arg Val Gly Pro Ser

•

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A)	LENGTH: 2415 base pairs
	TYPE: nucleic acid
	STRANDEDNESS: single
(D)	TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 3..2188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

	(X1)	SEQ	UENC	E DE	SCKI	PIIO	M: 2	EQ I	D NO							
CG A	TT T le S 1	CA T er P	TC C	TC G eu A	CT C la P 5	CC C	AC A lis A	GG T	cc c er L	TC T eu S 10	CC C	CA A	AA T ys T	ΆΤ Τ yr S	CC er 15	47
CAT His	CTT Leu	GTC Val	CTA Leu	GCC Ala 20	CAT His	CCC Pro	CCA Pro	GAC Asp	TAT Tyr 25	CTC Leu	AAG Lys	GAC Asp	CAG Gln	CTG Leu 30	TCC Ser	95
CCA Pro	CGC Arg	CCC Pro	CGA Arg 35	CCT Pro	CCA Pro	CTA Leu	GGC Gly	CTG Leu 40	TGC Cys	CAC His	CCG Pro	CTG Leu	CCT Pro 45	GCA Ala	GGA Gly	143
AGA Arg	CGC Arg	CCG Pro 50	GTC Val	CCG Pro	GGC Gly	CGG Arg	GTT Val 55	AGC Ser	CCC Pro	ATG Met	GGA Gly	ACG Thr 60	CAG Gln	CGC Arg	CTG Leu	191
TGT Cys	GGC Gly 65	CGC Arg	GGG Gly	ACT Thr	CAA Gln	GGC Gly 70	TGG Trp	CCT Pro	GGC Gly	TCA Ser	AGT Ser 75	GAA Glu	CAG Gln	CAC His	GTC Val	239
CAG Gln 80	GAG Glu.	GCG Ala	ACC Thr	TCG Ser	TCC Ser 85	GCG Ala	GGT Gly	TTG Leu	CAT His	TCT Ser 90	GGG Gly	GTG Val	GAC Asp	GAG Glu	CTG Leu 95	287
GGG Gly	GTT Val	CGG Arg	TCC Ser	GAG Glu 100	CCC Pro	GGT Gly	GGG Gly	AGG Arg	CTC Leu 105	CCG Pro	GAG Glu	CGC Arg	AGC Ser	CTG Leu 110	GGC Gly	335
CCA Pro	GCC Ala	CAC His	CCC Pro 115	GCG Ala	CCG Pro	GCG Ala	GCC Ala	ATG Met 120	GCA Ala	GGC Gly	ACC Thr	CTG Leu	GAC Asp 125	CTG Leu	GAC Asp	383
AAG Lys	GGC Gly	TGC Cys 130	ACG Thr	GTG Val	GAG Glu	GAG Glu	CTG Leu 135	CTC Leu	CGC Arg	GGG Gly	TGC Cys	ATC Ile 140	GAA Glu	GCC Ala	TTC Phe	431
GAT Asp	GAC Asp 145	Ser	GGG Gly	AAG Lys	GTG Val	CGG Arg 150	Asp	CCG Pro	CAG Gln	CTG Leu	GTG Val 155	Arg	ATG Met	TTC Phe	CTC Leu	479
ATG Met 160	Met	CAC His	CCC Pro	TGG Trp	TAC Tyr 165	ATC Ile	CCC Pro	TCC Ser	TCT Ser	CAG Gln 170	CTG Leu	GCG Ala	GCC Ala	AAG Lys	CTG Leu 175	527
CTC Leu	CAC His	ATC	TAC Tyr	CAA Gln 180	Gln	TCC Ser	CGG Arg	AAG Lys	GAC Asp 185	Asn	TCC Ser	AAT Asn	TCC Ser	CTG Leu 190	CAG Gln	579
GTG Val	AAA Lys	ACG Thr	TGC Cys 195	His	CTG Leu	GTC Val	AGG Arg	TAC Tyr 200	Trp	ATC Ile	TCC Ser	GCC Ala	TTC Phe 205	Pro	GCG Ala	623

						GAG Glu										671
GCT Ala	CTG Leu 225	CTA Leu	GAC Asp	CAA Gln	GAA Glu	GGG Gly 230	AAC Asn	CGA Arg	CGG Arg	CAC His	AGC Ser 235	AGC Ser	CTA Leu	ATC Ile	GAC Asp	719
						TAC Tyr										767
						AAG Lys										815
						GCG Ala										863
						TTT Phe										911
						CCC Pro 310										959
						GTG Val										1007
GCC Ala	CCG Pro	CAG Gln	CGG Arg	GCC Ala 340	CTG Leu	GTC Val	ATC Ile	ACA Thr	CAC His 345	TTT Phe	GTC Val	CAC His	GTG Val	GCG Ala 350	GAG Glu	1055
						AAC Asn										1103
GGC Gly	CTG Leu	AGC Ser 370	CAC His	AGC Ser	TCC Ser	ATC Ile	TCC Ser 375	CGC Arg	CTĆ Leu	AAG Lys	GAG Glu	ACC Thr 380	CAC His	AGC Ser	CAC His	1151
						AAG Lys 390										1199
ACG Thr 400	GCG Ala	ACA Thr	GGC Gly	AAC Asn	TAT Tyr 405	GGC Gly	AAC Asn	TAC Tyr	CGG Arg	CGT Arg 410	CGĢ Arg	CTG Leu	GCA Ala	GCC Ala	TGT Cys 415	1247
GTG Val	GGC Gly	TTC Phe	CGC Arg	TTC Phe 420	CCG Pro	ATC Ile	CTG Leu	GGT Gly	GTG Val 425	CAC His	CTC Leu	AAG Lys	GAC Asp	CTG Leu 430	GTG Val	1295
GCC Ala	CTG Leu	CAG Gln	CTG Leu 435	GCA Ala	CTG Leu	CCT Pro	GAC Asp	TGG Trp 440	CTG Leu	GAC Asp	CCA Pro	GCC Ala	CGG Arg 445	ACC Thr	CGG Arg	1343
CTC Leu	AAC Asn	GGG Gly 4 50	GCC Ala	AAG Lys	ATG Met	AAG Lys	CAG Gln 455	CTC Leu	TTT Phe	AGC Ser	ATC Ile	CTG Leu 460	GAG Glu	GAG Glu	CTG Leu	1391
GCC Ala	ATG Met 465	GTG Val	ACC Thr	AGC Ser	CTG Leu	CGG Arg 470	CCA Pro	CCA Pro	GTA Val	CAG Gln	GCC Ala 475	AAC Asn	CCC Pro	GAC Asp	CTG Leu	1439

480 485 490 495	
400	
CTG TAC CAG CTG TCC CTG CAG CGG GAG CCG CGC TCC AAG TCC TCG CCA Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro 500 505 510	35
ACC AGC CCC ACG AGT TGC ACC CCA CCA CCC CGG CCC CCG GTA CTG GAG Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu 515 520 525	83
GAG TGG ACC TCG GCT GCC AAA CCC AAG CTG GAT CAG GCC CTC GTG GTG Glu Trp Thr Ser Ala Ala Lys Pro Lys Leu Asp Gln Ala Leu Val Val 530 540	31
GAG CAC ATC GAG AAG ATG GTG GAG TCT GTG TTC CGG AAC TTT GAC GTC Glu His Ile Glu Lys Met Val Glu Ser Val Phe Arg Asn Phe Asp Val 545 550 555	79
GAT GGG GAT GGC CAC ATC TCA CAG GAA GAA TTC CAG ATC ATC CGT GGG Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe Gln Ile Ile Arg Gly 560 575	27
AAC TTC CCT TAC CTC AGC GCC TTT GGG GAC CTC GAC CAG AAC CAG GAT Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asn Gln Asp 580 585 590	75
GGC TGC ATC AGC AGG GAG GAG ATG GTT TCC TAT TTC CTG CGC TCC AGC Gly Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser 595 600 605	123
TCT GTG TTG GGG GGG CGC ATG GGC TTC GTA CAC AAC TTC CAG GAG AGC Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser 610 615 620	371
AAC TCC TTG CGC CCC GTC GCC TGC CGC CAC TGC AAA GCC CTG ATC CTG Asn Ser Leu Arg Pro Val Ala Cys Arg His Cys Lys Ala Leu Ile Leu 625 630 635	19
GGC ATC TAC AAG CAG GGC CTC AAA TGC CGA GCC TGT GGA GTG AAC TGC Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys 640 655 655	67
CAC AAG CAG TGC AAG GAT CGC CTG TCA GTT GAG TGT CGG CGC AGG GCC His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Ala 660 665 670)15
CAG AGT GTG AGC CTG GAG GGG TCT GCA CCC TCA CCC TCA CCC ATG CAC Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His 675 680 685	063
AGC CAC CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG Ser His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg 690 695 700	111
CGA GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GAG GTA CAG ACG GTG Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Val Gln Thr Val 705 710 715	159
GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG TGGTTGGATC 2 Glu Asp Gly Val Phe Asp Ile His Leu 720 725	208
AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA GCAGGGAGCC TGGGGGTGTC 2	268
GGGGCAGGAG GCTGGGGATG GGGGTGGGAT ATGAGGGTGG CATGCAGCTG AGGGCAGGGC	328

CAGGGCTGGT	GTCCCTAAGG	TTGTACAGAC	TCTTGTGAAT	ATTTGTATTT	TCCAGATGGA	2388
ATAAAAAGGC	CCGTGTAATT	AACCTTC		•		2415

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 728 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE-DESCRIPTION: SEQ ID NO:5:

Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr Ser His

1 5 10 15

Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln Leu Ser Pro
20 25 30

Arg Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu Pro Ala Gly Arg

Arg Pro Val Pro Gly Arg Val Ser Pro Met Gly Thr Gln Arg Leu Cys 50 60

Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser Ser Glu Gln His Val Gln 65 70 75 80

Glu Ala Thr Ser Ser Ala Gly Leu His Ser Gly Val Asp Glu Leu Gly 85 90 95

Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser Leu Gly Pro
100 105 110

Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp Leu Asp Lys
115 120 125

Gly Cys Thr Val Glu Glu Leu Leu Arg Gly Cys Ile Glu Ala Phe Asp 130 135 140

Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met 145 150 155 160

Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu 165 170 175

His Ile Tyr Gln Gln Ser Arg Lys Asp Asn Ser Asn Ser Leu Gln Val

Lys Thr Cys His Leu Val Arg Tyr Trp Ile Ser Ala Phe Pro Ala Glu 195 200 205

Phe Asp Leu Asn Pro Glu Leu Ala Glu Gln Ile Lys Glu Leu Lys Ala 210 220

Leu Leu Asp Gln Glu Gly Asn Arg Arg His Ser Ser Leu Ile Asp Ile 225 230 235 240

Asp Ser Val Pro Thr Tyr Lys Trp Lys Arg Gln Val Thr Gln Arg Asn 245 250 255

Pro Val Gly Gln Lys Lys Arg Lys Met Ser Leu Leu Phe Asp His Leu 260 265 270

Glu	Pro	Met 275	Glu	Leu	Ala	Glu	His 280	Leu	Thr	Tyr	Leu	Glu 285	Tyr	Arg	Ser
Phe	Cys 290	Lys	Ile	Leu	Phe	Gln 295	Asp	Tyr	His	Ser	Phe 300	Val	Thr	His	Gly
Cys 305	Thr	Val	Asp	Asn	Pro 310	Val	Leu	Glu	Arg	Phe 315	Ile	Ser	Leu	Phe	Asn 320
Ser	Val	Ser	Gln	Trp 325	Val	Gln	Leu	Met	Ile 330	Leu	Ser	Lys	Pro	Thr 335	Ala
Pro	Gln	Arg	Ala 340	Leu	Val	Ile	Thr	His 345	Phe	Val	His	Val	Ala 350	Glu	Lys
Leu	Leu	Gln 355	Leu	Gln	Asn	Phe	Asn 360	Thr	Leu	Met	Ala	Val 365	Val	Gly	Gly
Leu	Ser 370	His	Ser	Ser	Ile	Ser 375	Arg	Leu	Lys	Glu	Thr 380	His	Ser	His	Val
Ser 385	Pro	Glu	Thr	Ile	Lys 390	Leu	Trp	Glu	Gly	Leu 395	Thr	Glu	Leu	Val	Thr 400
Ala	Thr	Gly	Asn	Tyr 405	Gly	Asn	Tyr	Arg	Arg 410	Arg	Leu	Ala	Ala	Cys 415	Val
Gly	Phe	Arg	Phe 420	Pro	Ile	Leu	Gly	Val 425	His	Leu	Lys	Asp	Leu 430	Val	Ala
Leu	Gln	Leu 435	Ala	Leu	Pro	Asp	Trp 440	Leu	Asp	Pro	Ala	Arg 445	Thr	Arg	Leu
Asn	Gly 450	Ala	Lys	Met	Lys	Gln 455	Leu	Phe	Ser		Leu 460	Glu	Glu	Leu	Ala
Met 465	Val	Thr	Ser	Leu	Arg 470	Pro	Pro	Val	Gln	Ala 475	Asn	Pro	Asp	Leu	Leu 480
Ser	Leu	Leu	Thr	Val 485	Ser	Leu	Asp	Gln	Tyr 490	Gln	Thr	Glu	Asp	Glu 495	Leu
Tyr	Gln	Leu	Ser 500	Leu	Gln	Arg	Glu	Pro 505	Arg	Ser	Lys	Ser	Ser 510	Pro	Thr
Ser	Pro	Thr 515	Ser	Cys	Thr	Pro	Pro 520	Pro	Arg	Pro	Pro	Val 525	Leu	Glu	Glu
Trp	Thr 530	Ser	Ala	Ala	Lys	Pro 535	Lys	Leu	Asp	Gln	Ala 540	Leu	Val	Val	Glu
His 545	Ile	Glu	Lys	Met	Val 550	Glu	Ser	Val	Phe	Arg 555	Asn	Phe	Asp	Val	Asr 560
Gly	Asp	Gly	His	Ile 565	Ser	Gln	Glu	Glu	Phe 570		Ile	Ile	Arg	Gly 575	Asr
Phe	Pro	Tyr	Leu 580	Ser	Ala	Phe	Gly	Asp 585	Leu	Asp	Gln	Asn	Gln 590	Asp	Gly
Суѕ	Ile	Ser 595	Arg	Glu	Glu	Met	Val 600		Tyr	Phe	Leu	Arg 605	Ser	Ser	Ser
Val	Leu 610		Gly	Arg	Met	Gly 615		Val	His	Asn	Phe 620		Glu	Ser	Ası
Ser	Leu	Ara	Pro	Va1	Ala	Cvs	Ara	His	Cys	Lvs	Ala	Leu	Ile	Leu	Glv

529

625					630					635					640	
Ile	Tyr	Lys	Gln	Gly 645	Leu	Lys	Суѕ	Arg	Ala 650	Cys	Gly	Val	Asn	Cys 655	His	
Lys	Gln	Cys	Lys 660	qeA	Arg	Leu	Ser	Val 665	Glu	Cys	Arg	Arg	Arg 670	Ala	Gln	
Ser	Val	Ser 675	Leu	Glu	Gly	Ser	Ala 680	Pro	Ser	Pro	Ser	Pro 685	Met	His	Ser	
His	His 690	His	Arg	Ala	Phe	Ser 695	Phe	Ser	Leu	Pro	Arg 700	Pro	Gly	Arg	Arg	
Gly 705	Ser	Arg	Pro	Pro	Glu 710	Ile	Arg	Glu	Glu	Glu 715	Val	Gln	Thr	Val	Glu 720	
Asp	Gly	Val	Phe	Asp 725	Ile	His	Leu									
(2)	INFO	ORMAT	TON	FOR	SEO	ID N	IO : 6 :									
(2)			1011	. 01	220			•								
	(i)			CE CI ENGTI												
		(E	3) TY	PE:	nucl	leic	acid	1	. 5							
				POLC			_	,le								
			· .		•		aı									
	(ii)	MOI	LECUI	LE TY	PE:	DNA										
	(ix)	FE#													-	
			-	ME/F CATI			. 208	3								
			, <u> </u>													
	(xi)	SEC	OUENC	E DE	ESCRI	PTIC	N: S	EO I	D NO):6:						
CC 3/1			-								20020	maaa	יאמי כ	mmon	CCTAG	
														-		60
CCC	ATCCC	CC A	AGACI	ATCI	C A	IGGAC	CAGC	TGT	CCCC	ACG	cccc	CGAC	CT C	CACI	AGGCC	120
TGTC	CCAC	CC C	CTGC	CTGC	A GO	AAGA	CGCC	CGG	TCCC	GGG	CCGG	GTTA	GC C	CCAT	GGGAA	180
CGGG	GTTC	GG 7	CCGA	GCCC	G GI	GGGA	GGCI	ccc	GGAG	CGC	AGCC	TGGG	CC C	AGCC	CACCC	240
CGC	CCGG	CG C		TG G												289
	•		M	iet A	Ala G	Sly T	hr L	eu A 5	Asp I	eu A	sp L	ys G	10	lys T	hr	
GTG Val	GAG	GAG	CTG	CTC Leu	CGC	GGG	TGC	ATC	GAA	GCC	TTC	GAT	GAC	TCC	GGG	337
V u1	GIU	15	Бец	Бец	Arg	Gly	20	116	Giu	AIG	FIIE	25	ASP	ser	GIY	
AAG	GTG	CGG	GAC	CCG	CAG	CTG	GTG	CGC	ATG	TTC	СТС	ATG	ЭТА	CAC	CCC	385
Lys	Val	Arg	Asp	Pro	Gln	Leu	Val	Arg	Met	Phe	Leu	Met	Met	His	Pro	303
	30					35					40					
TGG	TAC	ATC	CCC	TCC	TCT	CAG	CTG	GCG	GCC	AAG	CTG	CTC	CAC	ATC	TAC	433
Trp 45	Tyr	Ile	Pro	Ser	Ser 50	Gln	Leu	Ala	Ala	Lys 55	Leu	Leu	His	Ile	Tyr 60	
CAA Gln	CAA Gln	TCC	CGG	AAG Lys	GAC	AAC Asp	TCC	AAT Asn	TCC	CTG	CAG	GTG Val	AAA	ACG	TGC	481
		J-G-1	ar 9	65					70	260	3111	401	пХЭ	75	CYS	

CAC CTG GTC AGG TAC TGG ATC TCC GCC TTC CCA GCG GAG TTT GAC TTG

His	Leu	Val	Arg 80	Tyr	Trp	Ile	Ser	Ala 85	Phe	Pro	Ala	Glu	Phe 90	Asp	Leu	
AAC Asn	CCG Pro	GAG Glu 95	TTG Leu	GCT Ala	GAG Glu	CAG Gln	ATC Ile 100	AAG Lys	GAG Glu	CTG Leu	AAG Lys	GCT Ala 105	CTG Leu	CTA Leu	GAC Asp	577
CAA Gln	GAA Glu 110	GGG Gly	AAC Asn	CGA Arg	CGG Arg	CAC His 115	AGC Ser	AGC Ser	CTA Leu	ATC Ile	GAC Asp 120	ATA Ile	GAC Asp	AGC Ser	GTC Val	625
CCT Pro 125	ACC Thr	TAC Tyr	AAG Lys	TGG Trp	AAG Lys 130	CGG Arg	CAG Gln	GTG Val	ACT Thr	CAG Gln 135	CGG Arg	AAC Asn	CCT Pro	GTG Val	GGA Gly 140	673
CAG Gln	AAA Lys	AAG Lys	CGC Arg	AAG Lys 145	ATG Met	TCC Ser	CTG Leu	TTG Leu	TTT Phe 150	GAC Asp	CAC His	CTG Leu	GAG Glu	CCC Pro 155	ATG Met	721
GAG Glu	CTG Leu	GCG Ala	GAG Glu 160	CAT His	CTC Leu	ACC Thr	TAC Tyr	TTG Leu 165	GAG Glu	TAT Tyr	CGC Arg	TCC Ser	TTC Phe 170	TGC Cys	AAG Lys	769
ATC Ile	CTG Leu	TTT Phe 175	CAG Gln	GAC Asp	TAT Tyr	CAC His	AGT Ser 180	TTC Phe	GTG Val	ACT Thr	CAT His	GGC Gly 185	TGC Cys	ACT Thr	GTG Val	817
GAC Asp	AAC Asn 190	CCC Pro	GTC Val	CTG Leu	GAG Glu	CGG Arg 195	TTC Phe	ATC Ile	TCC Ser	CTC Leu	TTC Phe 200	AAC Asn	AGC Ser	GTC Val	TCA Ser	865
CAG Gln 205	TGG Trp	GTG Val	CAG Gln	CTC Leu	ATG Met 210	ATC Ile	CTC Leu	AGC Ser	AAA Lys	CCC Pro 215	ACA Thr	GCC Ala	CCG Pro	CAG Gln	CGG Arg 220	913
GCC Ala	CTG Leu	GTC Val	ATC Ile	ACA Thr 225	CAC His	TTT Phe	GTC Val	CAC His	GTG Val 230	GCG Ala	GAG Glu	AAG Lys	CTG Leu	CTA Leu 235	CAG Gln	961
CTG Leu	CAG Gln	AAC Asn	TTC Phe 240	AAC Asn	ACG Thr	CTG Leu	ATG Met	GCA Ala 245	GTG Val	GTC Val	GGG Gly	GGC Gly	CTG Leu 250	AGC Ser	CAC His	1009
AGC Ser	TCC Ser	ATC Ile 255	TCC Ser	CGC Arg	CTC Leu	AAG Lys	GAG Glu 260	Thr	CAC His	AGC Ser	CAC His	GTT Val 265	AGC Ser	CCT Pro	GAG Glu	1057
ACC Thr	ATC Ile 270	Lys	CTC Leu	TGG Trp	GAG Glu	GGT Gly 275	Leu	ACG Thr	GAA Glu	CTA Leu	GTG Val 280	Thr	GCG Ala	ACA Thr	GGC Gly	1105
AAC Asn 285	Tyr	GGC Gly	AAC Asn	TAC Tyr	CGG Arg 290	CGT Arg	CGG Arg	CTG Leu	GCA Ala	GCC Ala 295	Cys	GTG Val	GGC Gly	TTC Phe	CGC Arg 300	1153
TTC Phe	CCG Pro	ATC Ile	CTG Leu	GGT Gly 305	Val	CAC His	CTC Leu	AAG Lys	GAC Asp 310	Leu	GTG Val	GCC Ala	CTG Leu	CAG Gln 315	CTG Leu	1201
GCA Ala	CTG Leu	CCT Pro	GAC Asp 320	Trp	CTG Leu	GAC Asp	CCA Pro	GCC Ala 325	Arg	ACC Thr	CGG Arg	CTC Leu	AAC Asn 330	GGG Gly	GCC Ala	1249
AAG Lys	ATG Met	AAG Lys 335	Gln	CTC Lev	TTT Phe	AGC Ser	Ile 340	Leu	GAG Glu	GAG Glu	CTG Leu	GCC Ala 345	Met	GTG Val	ACC	1297

		CAC His	TAA1	TAGAT	GC 1	rgtgo	TTGC	SA TO	CAAGO	SACTO	ATI	CCTC	CCT	2120
				GAG Glu 595										2065
				CTG Leu										2017
				TCA Ser										1969
				GAG Glu										1921
				GCC Ala										1873
				TGC Cys 515										1825
				CAC His										1777
				TAT Tyr										1729
				CTC Leu										1681
				TTC Phe										1633
				TTC Phe 435										1585
				GAT. Asp										1537
				CGG Arg										1489
				CGC Arg										1441
				TAT Tyr										1393
				CAG Gln 355									CTC Leu	1345

TTAACCTTC						2309
GGTTGTACAG	ACTCTTGTGA	ATATTTGTAT	TTTCCAGATG	GAATAAAAAG	GCCCGTGTAA	2300
TGGGGGTGGG	ATATGAGGGT	GGCATGCAGC	TGAGGGCAGG	GCCAGGGCTG	GTGTCCCTAA	2240
TGGAGAAAAT	ACTTCAACCA	GAGCAGGGAG	CCTGGGGGTG	TCGGGGCAGG	AGGCTGGGGA	2180

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 609 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Ala Gly Thr Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu 15

Leu Arg Gly Cys Ile Glu Ala Phe Asp Asp Ser Gly Lys Val Arg Asp Pro Gln Leu Val Arg Met Phe Leu Met Met His Pro Trp Tyr Ile Pro Ser Ser Gln Leu Ala Ala Lys Leu Leu His Ile Tyr Gln Gln Ser Arg 65

Lys Asp Asn Ser Asn Ser Leu Gln Val Lys Thr Cys His Leu Val Arg 80

Tyr Trp Ile Ser Ala Phe Pro Ala Glu Phe Asp Leu Asn Pro Glu Leu 85

Ala Glu Gln Ile Lys Glu Leu Lys Ala Leu Leu Asp Gln Glu Gly Asn 100

Arg Arg His Ser Ser Leu Ile Asp Ile Asp Ser Val Pro Thr Tyr Lys 110

Trp Lys Arg Gln Val Thr Gln Arg Asn Pro Val Gly Gln Lys Lys Arg 130

His Leu Thr Tyr Leu Glu Tyr Arg Ser Phe Cys Lys Ile Leu Phe Gln
165 170 175

Lys Met Ser Leu Leu Phe Asp His Leu Glu Pro Met Glu Leu Ala Glu

Asp Tyr His Ser Phe Val Thr His Gly Cys Thr Val Asp Asn Pro Val

Leu Glu Arg Phe Ile Ser Leu Phe Asn Ser Val Ser Gln Trp Val Gln
195 200 205

Leu Met Ile Leu Ser Lys Pro Thr Ala Pro Gln Arg Ala Leu Val Ile 210 215 220

Thr His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe 225 230 235 240

Asn Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser 245 250 255

Leu

Arg Leu Lys Glu Thr His Ser His Val Ser Pro Glu Thr Ile Lys Leu Trp Glu Gly Leu Thr Glu Leu Val Thr Ala Thr Gly Asn Tyr Gly Asn Tyr Arg Arg Arg Leu Ala Ala Cys Val Gly Phe Arg Phe Pro Ile Leu Gly Val His Leu Lys Asp Leu Val Ala Leu Gln Leu Ala Leu Pro Asp 310 Trp Leu Asp Pro Ala Arg Thr Arg Leu Asn Gly Ala Lys Met Lys Gln Leu Phe Ser Ile Leu Glu Glu Leu Ala Met Val Thr Ser Leu Arg Pro Pro Val Gln Ala Asn Pro Asp Leu Leu Ser Leu Leu Thr Val Ser Leu Asp Gln Tyr Gln Thr Glu Asp Glu Leu Tyr Gln Leu Ser Leu Gln Arg Glu Pro Arg Ser Lys Ser Ser Pro Thr Ser Pro Thr Ser Cys Thr Pro Pro Pro Arg Pro Pro Val Leu Glu Glu Trp Thr Ser Ala Ala Lys Pro 410 Lys Leu Asp Gln Ala Leu Val Val Glu His Ile Glu Lys Met Val Glu 420 Ser Val Phe Arg Asn Phe Asp Val Asp Gly Asp Gly His Ile Ser Gln 440 Glu Glu Phe Gln Ile Ile Arg Gly Asn Phe Pro Tyr Leu Ser Ala Phe Gly Asp Leu Asp Gln Asp Gln Asp Gly Cys Ile Ser Arg Glu Glu Met Val Ser Tyr Phe Leu Arg Ser Ser Ser Val Leu Gly Gly Arg Met Gly Phe Val His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys 505 Arg His Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu Ser Val Glu Cys Arg Arg Arg Ala Gln Ser Val Ser Leu Glu Gly Ser Ala Pro Ser Pro Ser Pro Met His Ser His His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg Arg Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr Val Glu Asp Gly Val Phe Asp Ile His 600

- 72 -

(2) INFORMATION	FOR	SEQ	ID	NO:8:
-----------------	-----	-----	----	-------

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 832 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 11..733

ENCE DESCRIPTION: SEO ID NO:8:

	(xi)	SEQ	UENC	E DE	SCRI	PTIO	N: S	EQ I	D NO	:8:						
GCCC	GCCG	CC A	TG C et P	CG C ro P	CC T ro L	TA C eu L	TG C eu P 5	cc c	TG C eu A	GC C	TG T eu C	GC C ys A 10	GG C rg L	TG T eu T	GG 'rp	49
CCC Pro	CGC Arg 15	AAC Asn	CCT Pro	CCC Pro	TCC Ser	CGG Arg 20	CTC Leu	CTC Leu	GGA Gly	GCG Ala	GCC Ala 25	GCC Ala	GGG Gly	CAG Gln	CGG Arg	97
TCC Ser 30	AGA Arg	CCC Pro	AGT Ser	ACT Thr	TAT Tyr 35	TAT Tyr	GAA Glu	CTG Leu	TTG Leu	GGG Gly 40	GTG Val	CAT His	CCT Pro	GGT Gly	GCC Ala 45	145
AGC Ser	ACT Thr	GAG Glu	GAA Glu	GTT Val 50	AAA Lys	CGA Arg	GCT Ala	TTC Phe	TTC Phe 55	TCC Ser	AAG Lys	TCC Ser	AAA Lys	GAG Glu 60	CTG Leu	193
CAC His	CCA Pro	GAC Asp	CGG Arg 65	GAC Asp	CCT Pro	GGG Gly	AAC Asn	CCA Pro 70	AGC Ser	CTG Leu	CAC His	AGC Ser	CGC Arg 75	TTT Phe	GTG Val	241
GAG Glu	CTG Leu	AGC Ser 80	GAG Glu	GCA Ala	TAC Tyr	CGT Arg	GTG Val 85	CTC Leu	AGC Ser	CGT Arg	GAG Glu	CAG Gln 90	AGC Ser	CGC Arg	CGC Arg	289
AGC Ser	TAT Tyr 95	GAT Asp	GAC Asp	CAG Gln	CTC Leu	CGC Arg 100	TCA Ser	GGT Gly	AGT Ser	CCC Pro	CCA Pro 105	AAG Lys	TCT Ser	CCA Pro	CGA Arg	337
ACC Thr 110	Thr	GTC Val	CAT His	GAC Asp	AAG Lys 115	TCT Ser	GCC Ala	CAC His	CAA Gln	ACA Thr 120	CAC His	AGC Ser	TCC Ser	TGG Trp	ACA Thr 125	385
CCC Pro	CCC Pro	AAC Asn	GCA Ala	CAG Gln 130	TAC Tyr	TGG Trp	TCC Ser	CAG Gln	TTT Phe 135	His	AGC Ser	GTG Val	AGG Arg	CCA Pro 140	CAG Gln	433
GGG Gly	CCC	CAG Gln	TTG Leu 145	AGG Arg	CAG Gln	CAG Gln	CAA Gln	CAC His 150	Lys	CAA Gln	AAC Asn	AAA Lys	CAA Gln 155	GTG Val	CTG Leu	481
GGG	TAC Tyr	TGC Cys 160	Leu	CTC Leu	CTC Leu	ATG Met	CTG Leu 165	Ala	GGC	ATG Met	GGC Gly	CTG Leu 170	His	TAC Tyr	ATT Ile	529
GCC Ala	TTC Phe 175	Arg	AAG Lys	GTG Val	AAG Lys	CAG Gln 180	Met	CAC	CTT Leu	AAC Asn	TTC Phe 185	Met	GAT Asp	GAA Glu	AAG Lys	577

																	•
GAT Asp 190	CGG Arg	ATC Ile	ATC Ile	ACA Thr	GCC Ala 195	TTC Phe	TAC Tyr	AAC Asn	GAA Glu	GCC Ala 200	CGG Arg	GCA Ala	CGG Arg	GCC Ala	AGG Arg 205		625
GCC Ala	AAC Asn	AGA Arg	GGC Gly	ATC Ile 210	CTT Leu	CAG Gln	CAG Gln	GAG Glu	CGA Arg 215	CAA Gln	CGG Arg	CTA Leu	GGG Gly	CAG Gln 220	CGG Arg		673
CAG Gln	CCG Pro	CCA Pro	CCA Pro 225	TCC Ser	GAG Glu	CCA Pro	ACC Thr	CAA Gln 230	GGC Gly	CCC Pro	GAG Glu	ATC Ile	GTG Val 235	CCC Pro	CGG Arg		721
	GCC Ala			TGA	GGGG	SCTC	ACC!	rgga'	rgg (GCC1	rgca(GT G	CGTT	CCCG	C		773
TTTC	CTT	CCT 1	rccc	rgga	CG GC	CCG	CTCC	C CG	AAAC	GCGC	GCA	ATAA	AGT (GATT	CGCAG		832
(2)	INFO	ORMA'	rion	FOR	SEQ	ID 1	10:9	:									
	ı	(i) :	(A)	LEI TYI	CHAR NGTH: PE: & POLOG	241	l ami	ino a id	: acids	· •							
	(:	ii) 1	MOLEC	CULE	TYPE	E: pı	otei	in									
	()	ci) S	SEQUE	ENCE	DESC	RIP	NOI!	: SE(Q ID	ио:9	€:		•	•			
Met 1	Pro	Pro	Leu	Leu 5	Pro	Leu	Arg	Leu	Cys 10	Arg	Leu	Trp	Pro	Arg 15	Asn		
Pro	Pro	Ser	Arg 20	Leu	Leu	Gly	Ala	Ala 25	Ala	Gly	Gln	Arg	Ser 30	Arg	Pro		
Ser	Thr	Tyr 35	Tyr	Glu	Leu	Leu	Gly 40	Val	His	Pro	Gly	Ala 45	Ser	Thr	Glu		
Glu	Val 50	Lys	Arg	Ala	Phe	Phe 55	Ser	Lys	Ser	Lys	Glu 60	Leu	His	Pro	Asp		
Arg 65	Asp	Pro	Gly	Asn	Pro 70	Ser	Leu	His	Ser	Arg 75	Phe	Val	Glu	Leu	Ser 80		
Glu	Ala	Tyr	Arg	Val 85	Leu	Ser	Arg	Glu	Gln 90	Ser	Arg	Arg	Ser	Tyr 95	Asp		
Asp	Gln	Leu	Arg 100	Ser	Gly	Ser	Pro	Pro 105	Lys	Ser	Pro	Arg	Thr 110	Thr	Val	•	
His	Asp	Lys 115	Ser	Ala	His	Gln	Thr 120	His	Ser	Ser	Trp	Thr 125	Pro	Pro	Asn		
Ala	Gln 130	Tyr	Trp	Ser	Gln	Phe 135	His	Ser	Val	Arg	Pro 140	Gln	Gly	Pro	Gln		
Leu 145	Arg	Gln	Gln	Gln	His 150	Lys	Gln	Asn	Lys	Gln 155	Val	Leu	Gly	Tyr	Cys 160		
Leu	Leu	Leu	Met	Leu 165	Ala	Gly	Met	Gly	Leu 170	His	Tyr	Ile	Ala	Phe 175	Arg		
Lys	Val	Lys	Gln 180	Met	His	Leu	Asn	Phe	Met	Asp	Glu	Lys	Asp	Arg	Ile		

Ile Thr Ala Phe Tyr Asn Glu Ala Arg Ala Arg Ala Arg Ala Asn Arg

PCT/AU98/00380

- 74 -

205 200 195

Gly Ile Leu Gln Gln Glu Arg Gln Arg Leu Gly Gln Arg Gln Pro Pro 215

Pro Ser Glu Pro Thr Gln Gly Pro Glu Ile Val Pro Arg Gly Ala Gly 230 235

Pro

SEQ ID Nos: 10-18 25-36

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 300 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (ix) FEATURE:

 - (A) NAME/KEY: CDS
 (B) LOCATION: 170..300
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGATTTCATT CCTCGCTCCC	CACAGGTCCC	TCTCCCCAAA	ATATTCCCAT	CTTGTCCTAG	60

CCCATCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG CCCCGACCT CCACTAGGCC 120

TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG CCGGGTTAG CCC CAT 175 Pro His

GGG AAC GGG GTT CGG TCC GAG CCC GGT GGG AGG CTC CCG GAG CGC AGC Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu Arg Ser

CTG GGC CCA GCC CAC CCC GCG CCG GCG GCC ATG GCA GGC ACC CTG GAC 271 Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr Leu Asp

CTG GAC AAG GGC TGC ACG GTG GAG GAG CT 300 Leu Asp Lys Gly Cys Thr Val Glu Glu Leu 40

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

WO 98/53061 PCT/AU98/00380

15

- 75 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Pro His Gly Asn Gly Val Arg Ser Glu Pro Gly Gly Arg Leu Pro Glu

Arg Ser Leu Gly Pro Ala His Pro Ala Pro Ala Ala Met Ala Gly Thr 25

Leu Asp Leu Asp Lys Gly Cys Thr Val Glu Glu Leu

- (2) INFORMATION FOR SEQ ID NO:9:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9: GGGATCCCCC TGGTC
- (2) INFORMATION FOR SEQ ID NO:10:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Asp Val Asp Glu Glu Asp Glu Val Glu Asp Ile Glu Phe

- (2) INFORMATION FOR SEQ ID NO:11:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Asp Val Asp Gly Asp Gly His Ile Ser Gln Glu Glu Phe

- (2) INFORMATION FOR SEQ ID NO:12:
 - (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 13 amino acids

PCT/AU98/00380 WO 98/53061

- 76 -

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Peptide
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asp His Asp Arg Asp Gly Phe Ile Ser Gln Glu Glu Phe

- (2) INFORMATION FOR SEQ ID NO:13:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Asp Gln Asn Gln Asp Gly Cys Ile Ser Arg Glu Glu Met

- (2) INFORMATION FOR SEQ ID NO:14:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 13 amino acids (B) TYPE: amino acid

 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asp Val Asp Met Asp Gly Gln Ile Ser Lys Asp Glu Leu 10

- (2) INFORMATION FOR SEQ ID NO:15:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids

 - (B) TYPE: amino acid(C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

His Phe Val His Val Ala Glu Lys Leu Leu Gln Leu Gln Asn Phe Asn

Thr Leu Met Ala Val Val Gly Gly Leu Ser His Ser Ser Ile Ser Arg

PCT/AU98/00380 WO 98/53061

- 77 -

Leu Lys Glu Thr His 35

- (2) INFORMATION FOR SEQ ID NO:16:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Lys Phe Val His Val Ala Lys His Leu Arg Lys Ile Asn Asn Phe Asn

Thr Leu Met Ser Val Val Gly Gly Ile Thr His Ser Ser Val Ala Arg 25

Leu Ala Lys Thr Tyr

- (2) INFORMATION FOR SEQ ID NO:17:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

His Asn Phe Gln Glu Ser Asn Ser Leu Arg Pro Val Ala Cys Arg His

Cys Lys Ala Leu Ile Leu Gly Ile Tyr Lys Gln Gly Leu Lys Cys Arg

Ala Cys Gly Val Asn Cys His Lys Gln Cys Lys Asp Arg Leu Ser Val

Glu Cys 50

- (2) INFORMATION FOR SEQ ID NO:18:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: Peptide
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

His Asn Phe His Glu Thr Thr Phe Leu Thr Pro Thr Thr Cys Asn His

	1				5					10					15		
	Cys	Asn	Lys	Leu 20	Leu	Trp	Gly	Ile	Leu 25	Arg	Gln	Gly	Phe	Lys 30	Cys	Lys	
٠	Asp	Cys	Gly 35	Leu	Ala	Val	His	Ser 40	Cys	Cys	Lys	Ser	Asn 45	Ala	Val	Ala	
	Glu	Cys 50															
(2)	INFO	RMATI	ON I	FOR :	SEQ :	ID N	0:19	:									
	(i)	(A) (B) (C)	LEI TYI	NGTH PE: 1 RAND	: 15 nucl	base eic a SS:	sing:	irs									
	(ii)	MOLI	ECULI	E TY	PE: 1	ONA											
	(xi)	SEQU	JENC	E DE	SCRI	PTIO	N: SI	EQ II	ON C	:19:							
GGGA	TCCC	CC TO	GTC														15
(2)	INFO	RMAT	ION I	FOR	SEQ :	ID N	0:20	:									
	(i)	(A) (B) (C)	LEI TY:	NGTH PE: 1 RAND	: 21 nucl	base eic SS:	sing:	irs									
	(ii)	MOLI	ECUL	Е ТҮ	PE: 1	DNA											
	(xi)	SEQ	JENC:	E DE	SCRI	PTIO	N: S	EQ I	D NO	:20:							
gaai	TCGG	CA CO	GAGC	CGAC	G G												21
(2)	INFO	RMAT:	ION :	FOR	SEQ	ID N	0:21	:									
	(i)	(A (B (C) LE) TY) ST	NGTH PE: RAND	: 78	bas eic SS:	STIC e pa acid sing ar	irs									
	(ii)	MOL	ECUL	Е ТҮ	PE:	DNA											
	(xi)	SEQ	UENC	E DE	SCRI	PTIO	N: S	EQ I	ои о	:21:							
ATG	GAGCA	GA A	GCTG	ATCT	C CG	AGGA	GGAC	CTG	cccg	GGG	CAGC	TGGA	TC C	GCAG	CCA	C	60
ccc	GCGCC	GG C	GGCC	ATG													78
(2)	INFO	RMAT:	ION	FOR	SEQ	ID N	0:22	:									

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Pro Gly Ala Ala Gly

Ser Ala Ala His Pro Ala Pro Ala Ala Met 20

- (2) INFORMATION FOR SEQ ID NO:23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23: GGATCCGCAG CCCACCCCGC GCCGGCGGCC ATG

33

- (2) INFORMATION FOR SEQ ID NO:24:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 amino acids

 - (B) TYPE: amino acid (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
- Gly Ser Ala Ala His Pro Ala Pro Ala Ala Met
- (2) INFORMATION FOR SEQ ID NO:25:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA

20

GTCATAGTCT GTCTCCTACT

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
GGAC	AAAGTG TGTGATGAAC C	21
(2)	INFORMATION FOR SEQ ID NO:26:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
CTCA	TCCTCC GTCTGATACT G	21
(2)	INFORMATION FOR SEQ ID NO:27:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
GTAG	ATGTGG ATCAGCTTGG	20
(2)	INFORMATION FOR SEQ ID NO:28:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
AGGT	CGGAGAA TGGTCAAGG	19
(2)	INFORMATION FOR SEQ ID NO:29:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	

(2)	INFORMATION FOR SEQ ID NO:30:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
•	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:	
ACA!	PAGACAG CGTGCCTACC	20
(2)	INFORMATION FOR SEQ ID NO:31:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:	
TAC	AACCTTA GGGACACCAG	20
(2)	INFORMATION FOR SEQ ID NO:32:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	-
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
TGC	TGAGCCT GCTCACGGTG	20
(2)	INFORMATION FOR SEQ ID NO:33:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
CAA	GTGAACA GCACGTCC	18
(2)	INFORMATION FOR SEQ ID NO:34:	

(i) SEQUENCE CHARACTERISTICS:

PCT/AU98/00380

	(A) LENGTH: 21 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
((ii) MOLECULE TYPE: DNA	
((xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
GACTA	ATCTCA AGGACCAGCT G	21
(2) 1	INFORMATION FOR SEQ ID NO:35:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
((ii) MOLECULE TYPE: DNA	
,	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
GGTT	CGGTCC GAGCCCGG	18
(2)	INFORMATION FOR SEQ ID NO:36:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:	
	CGATAC TCCAAGTAGG T	21
GGAGG	CONTAC ICCANGIAGG I	21
(2)	INFORMATION FOR SEQ ID NO:37:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: DNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
AGCG	GGCCAG GCCCCTTC	18

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

(2) INFORMATION FOR SEQ ID NO:42:

	(A) LENGTH: 20 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear				
	(ii) MOLECULE TYPE: DNA				
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:				
CATO	CTGGTC CAATGCGCTC	20			
(2)	INFORMATION FOR SEQ ID NO:39:				
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
	(ii) MOLECULE TYPE: DNA				
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:				
GCAC	TGAGGA AGTTAAACGA GC	22			
(2)) INFORMATION FOR SEQ ID NO:40:				
٠	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
	(ii) MOLECULE TYPE: DNA				
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:				
GCTC	GTTTAA CTTCCTCAGT GC	22			
(2)	INFORMATION FOR SEQ ID NO:41:				
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
	(ii) MOLECULE TYPE: DNA				
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:				
CCTC	ACCURCE ACADACCCCC III				

		EQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
	(ii) M	IOLECULE TYPE: DNA			
	(xi) S	SEQUENCE DESCRIPTION: SEQ ID NO:42:			
ACCA	ACCAGCTCCG CTCAGGTAG 1				
(2)	INFORM	MATION FOR SEQ ID NO:43:			
	(i) S	SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
	(ii) M	MOLECULE TYPE: DNA			
	(xi) S	SEQUENCE DESCRIPTION: SEQ ID NO:43:			
TCCA	GGAGCT	r GTGTGTTTGG	20		
(2)	INFORM	MATION FOR SEQ ID NO:44:			
	(i) S	SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
	(ii) N	MOLECULE TYPE: DNA			
	, ,				
	(xi) 5	SEQUENCE DESCRIPTION: SEQ ID NO:44:			
CCAC	GTTTCA	C AGCGTGAGG	19		
(2)	INFOR	MATION FOR SEQ ID NO:45:			
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
	(ii) :	MOLECULE TYPE: DNA			
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:45:			

WO 98/53061

- 85 -

CLAIMS:

- 1. An isolated nucleic acid molecule comprising a sequence of nucleotides encoding or complementary to a sequence encoding an amino acid sequence having homology to a regulator of gene expression or a derivative of said gene regulator.
- 2. An isolated nucleic acid molecule according to claim 1 wherein the regulator comprises a zinc finger domain of an (HC₃)₂ type.
- 3. An isolated nucleic acid molecule according to claim 2 wherein the sequence of nucleotides or complementary sequence of nucleotides is selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3; (ii)
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- a nucleotide sequence capable of hybridising under low stringency conditions to the (iv) nucleotide sequence set forth in (i), (ii) or (iii).
- 4. An isolated nucleic acid molecule according to claim 1 wherein said gene regulator is a guanine nucleotide exchange factor (GEF) or a derivative thereof.
- 5. An isolated nucleic acid molecule according to claim 4 wherein the sequence of nucleotides is selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or (ii) 7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the

nucleotide sequence set forth in (i), (ii) or (iii).

- 6. An isolated nucleic acid molecule according to claim 1, wherein said gene regulator is a heat shock protein or is a heat shock binding protein or a derivative thereof.
- 7. An isolated nucleic acid molecule according to claim 6, wherein the sequence of nucleotides is selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
- 8. A genetic construct comprising a vector portion and a gene portion comprising a regulator of gene expression or a derivative thereof.
- 9. A genetic construct according to claim 8 wherein the gene portion comprises a zinc finger domain of (HC₃)₂ type.
- 10. A genetic construct according to claim 9 wherein the gene portion comprises a nucleotide sequence selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:2;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:3;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).

- 11. A genetic construct according to claim 8 wherein said gene portion is a nucleotide exchange factor (GEF) or derivative thereof.
- 12. A genetic construct according to claim 11 wherein the gene portion comprises a nucleotide sequence selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:4 or 6;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEQ ID NO:5 or7;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
- 13. A genetic construct according to claim 8 wherein the gene portion is a heat shock protein or a derivative thereof or a heat shock binding protein or derivative thereof.
- 14. A genetic construct according to claim 13 wherein the gene portion comprises a nucleotide sequence selected from:
- (i) a nucleotide sequence set forth in SEQ ID NO:8;
- (ii) a nucleotide sequence encoding an amino acid sequence set forth in SEO ID NO:9;
- (iii) a nucleotide sequence having at least about 40% similarity to the nucleotide sequence of (i) or (ii); and
- (iv) a nucleotide sequence capable of hybridising under low stringency conditions to the nucleotide sequence set forth in (i), (ii) or (iii).
- 15. A nucleic acid molecule encoding a gene regulator having the identifying characteristics of a molecule selected from MCG4, MCG7 and MCG18 having respective amino acid sequences of SEQ ID NO:3, SEQ ID NO:5 or 7 and SEQ ID NO:9.

- 16. A method of detecting a condition caused or facilitated by an aberration in mcg4, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg4 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.
- 17. A method of detecting a condition caused or facilitated by an aberration in *mcg4*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG4 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.
- 18. A method for detecting MCG4 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG4 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG4 complex to form, and then detecting said complex.
- 19. A method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg7 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.
- 20. A method of detecting a condition caused or facilitated by an aberration in mcg7, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG7 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.
- A method for detecting MCG7 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG7 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG7 complex to form, and then detecting said complex.

- A method of detecting a condition caused or facilitated by an aberration in mcg18, said method comprising determining the presence of a single or multiple nucleotide substitution, deletion and/or addition or other aberration to one or both alleles of said mcg18 wherein the presence of such a nucleotide substitution, deletion and/or addition or other aberration may be indicative of said condition or a propensity to develop said condition.
- A method of detecting a condition caused or facilitated by an aberration in *mcg18*, said method comprising screening for a single or multiple amino acid substitution, deletion and/or addition to MCG18 wherein the presence of such a mutation is indicative of or a propensity to develop said condition.
- 24. A method for detecting MCG18 or a derivative thereof in a biological sample said method comprising contacting said biological sample with an antibody specific for MCG18 or its derivatives or homologues for a time and under conditions sufficient for an antibody-MCG18 complex to form, and then detecting said complex.

FIG 1 (II)

FIG 1 (III)

FIG 1 (IV)

FIG 1 (V)

FIG 1

L	r gu CAG'	RE 1 Taa?	FIGURE 1 (I) TCAGTAAACA CAGA	CAGAC	3ACT(GACTGG GGATCGATC	BATCO	BATC	ATG Met 1	GGG G1y	CTT Leu	TGT Cys	AAG Lys 5	TGC Cys	CCC Pro	50
\[\frac{1}{2} \]	AAG L Lys	AGA Arg	AAG Lys 10	AGA AAG GTG Arg Lys Val	ACC Thr	AAC Asn	CTG	TTC Phe 15	TGC	TTC	GAA Glu	CAT His	CGG Arg 20	GTC Val	AAC Asn	95
<u></u> დ გ	GTC ' Val	TGC Cys	GAG Glu 25	CAC His	TGC	CTG	GTA Val	GCC Ala 30	AAT Asn	CAC His	GCC Ala	AAG Lys	TGC Cys 35	ATC Ile	GTC Val	140
ပပ	CAG Gln	TCC Ser	TAC Tyr 40	CTG	CAA Gln	TGG	CTC	CAA Gln 45	GAT	AGC Ser	GAC Asp	TAC Tyr	AAC Asn 50	CCC	AAT Asn	185
Ĕΰ	TGC	CGC Arg	CTG Leu 55	TGC	AAC Asn	ATA Ile	CCC	CTG Leu 60	GCC Ala	AGC Ser	CGA	GAG Glu	ACG Thr 65	ACC Thr	CGC Arg	230
ÜÄ	CTT	GTC Val	TGC Cys 70	\mathtt{TAT}	GAT	CTC	TTT Phe	CAC His	TGG Trp	GCC Ala	TGC	CTC	AAT Asn 80	GAA	CGT Arg	275

320	365	410	455	200	545
TGC	GGC	TGG	AGC	TCT	AGC
Cys	Gly	Trp	Ser	Ser	
CAG	GCT	AAC	GTG	TGG	GAC
Gln	Ala	Asn	Val	Trp	Asp
TAT	CTG	GTC	GTG	GAC	GTA
Tyr	Leu	Val	Val	ASP	Val
95	110	125	140	155	170
GGC	AAC	ACA	GAG	TCT	GAG
Gly	Asn	Thr	Glu	Ser	Glu
GCC	ACC	GCC	GAT	TTC	GAG
Ala	Thr	Ala	Asp	Phe	Glu
CCT Pro	CCA	CTG	ATC Ile	GAC Asp	CCA Pro
GCA	CCC	AAG	CTG	TCT	GGA
Ala	Pro	Lys	Leu	Ser	Gly
ACG	TTC	GAG	CCT	ACG	CCT
Thr	Phe	Glu	Pro	Thr	Pro
90	105	120	135	150	165
AAC	ATC	AGA	CTC	AAC	ACC
Asn	Ile	Arg		Asn	Thr
CGA	CCC	CTG	GGC	CTC	AGT
Arg	Pro		Gly	Leu	Ser
CCC	GGC	GCA	CTG	CCC	AGC
Pro	Gly	Ala	Leu		Ser
CTA CTA Leu	AAT Asn	TCC Ser	GGA Gly	GAG Glu	GCC Ala
CAG CAB	TGC	GCC	GCA	CCC	AAT
	Cys	Ala	Ala	Pro	Asn
	100	115	130	145	160
	AGC	GTG	CGG	GAG	TTT
	Ser	Val	Arg	Glu	Phe
FIGURE 1 GCT GCC Ala Ala	CCC	CCC	GCC Ala	CCA Pro	AGT Ser
		Substitute Sh	eet (Rule 26)		

_		10	0	S	0	Ω
590))	635	089	725	770	81
۵	Pro	GGC Gly	ACG Thr	GAT	CTG	CTC Leu
ري	Pro	ATG Met	GAT Asp	GAC Asp	CGG	CTG
נילט.	Arg 185	CAC His 200	TAT Tyr 215	TGT Cys 230	GCC Ala 245	ACC Thr 260
ري		ATC Ile	GTG Val	GAC Asp	CTG	CTG
ن	Ala	GTG Val	AAG Lys	GGA Gly	TGG Trp	CCG Pro
۵ ا		ACA Thr	AGG Arg	CAT His	GGT Gly	CGG Arg
ر ا		CAC His	CCT	CTC	TTG Leu	AAG Lys
ر ا ا		CAG Gln 195	GCC Ala 210	GGC G1y 225	GCC Ala 240	CGG Arg 255
Į.		GAG Glu	CAC His	CCA	CCG	TCT Ser
נ	Ala	CCC Pro	ACT Thr	ACA Thr	CGG	$^{ m GGG}_{ m G1Y}$
	Pro	CGG Arg	TTG Leu	CGG Arg	CGT Arg	GCT Ala
(H)	Ala	GGC Gly	CCC	GAC Asp	CGA Arg	CGG Arg
1 (II	Ala A. 175	CCA Pro 190	GAG G1u 205	GAT Asp 220	TAC Tyr 235	AGC Ser 250
		TCC	CCT Pro	GAT Asp	AAG	AGG Arg
FIGU	Ala Ser	GCT Ala	AAT Asn	CGG Arg	GAC	CTA
			Substitute Sheet	(Rule 26)	•	

860	905	950	1002	1052	1102	1152
FIGURE 1 (IV) CAG CGG GGG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG CAG CGG GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG CAG CGG GCG GGG CTG CTG CTA CTC TTG GGA CTG CTG GGC TTC CTG CAG CGG GCG CTG CTG CTA CTC TTG GGA CTG CTG CAG CAG CGG GGG CTG CTG CTA CTG CTG CAG CAG CGG GGG CTG CTG CTA CTG CAG CAG CTG CTG CTG CTG CTG CTG CAG CAG CGG GGG CTG CTG CTG CTG CAG CAG CTG CTG CTG CTG CTG CTG CAG CTG CTG CTG CTG CTG CTG CAG CTG CTG CTG CTG CTG CTG CTG CAG CTG CTG CTG CTG CTG CTG CTG CAG CTG CTG CTG CTG CTG CTG CTG CTG CTG CT	GCC CTC CTT GCC CTC ATG TCT CGC CTA GGC CGG GCC GCA GCT GAC Ala Leu Leu Ala Leu Met Ser Arg Leu Gly Arg Ala Ala Asp 280	AGC GAT CCC AAC CTG GAC CCA CTC ATG AAC CCT CAC ATC CGC GTG Ser Asp Pro Asn Leu Asp Pro Leu Met Asn Pro His Ile Arg Val	GGC CCC TCC TGA GCCCCCTTGC TTGTGGCTAG GCCAGCCTAG GATGTGGGTT S Gly Pro Ser * 310	CTGTGGAGGA GAGGCGGGGT AATGGGGAGG CTGAGGGCCAC CTCTTCACTG	CCCCTCTCCC TCAAGCCTAA GACACTAAGA CCCCAGACCC AAAGCCAAGT	CCACCAGAGT GGCTCGCAGG CCAGGCCTGG AGTCCCCGTG GGTCAAGCAT

GURE 1 (V) GTCTTGAC TTGC	TTTCTC	CCGGGTCTCC	RECTITICIC CCGGGICICC AGCCICCGAC CCCICGCCC	CCCTCGCCCC	1202
TGAAGGAGC TGGC	AGGTGG	GCAGGTGG AAATAAACAA CAACTTTATT	CAACTTTATT		1242

9

gb|AA155210|AA155210 mr98e01.rl Stratagene mouse embryonic carcinoma Ω (#937317) Mus musculus cDNA clone 605496

FIGURE

277 MGLCKCPKRKVTNLFCFEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCNTPL MGLCKCPKRKVTNLFCFEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCN PL 98 Sbjct:

MGLCKCPKRKVTNLFCFEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCNIPL

Query:

109

single read end, elegans cDNA clone yklllg3:5' CELK111G3F dbj | b75913 |

FIGURE

99 7 PKRKVTNLFCFEHRVNVCEHCLVANHAKCIVQSYLQWLQDSDYNPNCRLCNIPLASRETT C+VQSYL WL D DY+PNC LC LV NH PKRKVTNLF +EHRVNVCE Query:

180 PKRKVTNLFXYEHRVNVCELXLVDNHPNCVVQSYLTWLTDQDYDPNCSLCKTTLXEGDTI Sbjct:

PSCNGPIFPPNQ +FPP+Q P C+ 98 RLVCYDLFHWACLNERAAQLPRNTAPAGYQCP 98 TAP GY+CP Д L HW C +E RL C 29 Query:

310 PCCSQEVFPPDQ 275 276 RLNCLHLLHWKCFDEWXGNFPDTTAPXGYRCP 181 Sbjct:

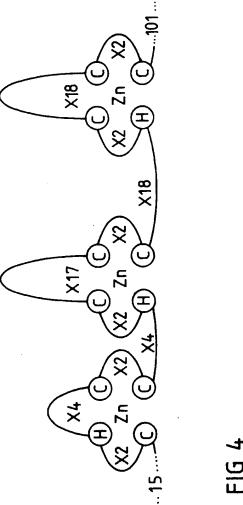


FIGURE 5

gi|500728 (U10402) C34E10.5 gene product 100 YLBS_CAEEL HYPOTHETICAL 146.8 KD PROTEIN C34E10.5 CNIPLASRETTRLVCYDLFHWACLNERAAQLPRNTAPAGYQCPSC ø LF W C+ E [Caenorhabditis elegans] CHROMOSOME III + L C C+I I ++ sp | P46580 56 Query:

CSICLENKNPSALFCGHLFCWTCIQEHAVAATSSASTSSARCPQC Sbjct:1222

1266

500

Д

221

11/85

(L29051) homologous to GATA-binding transcription factor 703468 gi

FIGURE

[Schizosaccharomyces pombe]

58 CIVQSYLQWLQDSDYNPNCRLCNI 35 Query:

NP C Q+ M+ + ပ

198 CATINTPKWRRDESGNPICNACGL SSTPGPEEVDSASAAPAFYSQAPRPPASPGRPEQHTVIHMGNPEPLTHAPRKVYDTRDDD **>**+ 162 Query:

ASLLINPEEPPSNSDKQPSMSNGPKSEVSPSQSQQAPLIQSSTSPVSLQFPPEVQGSNVDK SP+ +Q Б Ŋ Ø Ŋ PEE ჯ +

RTPGLH 227 222 Query:

441

Sbjct:

RNYALN 506

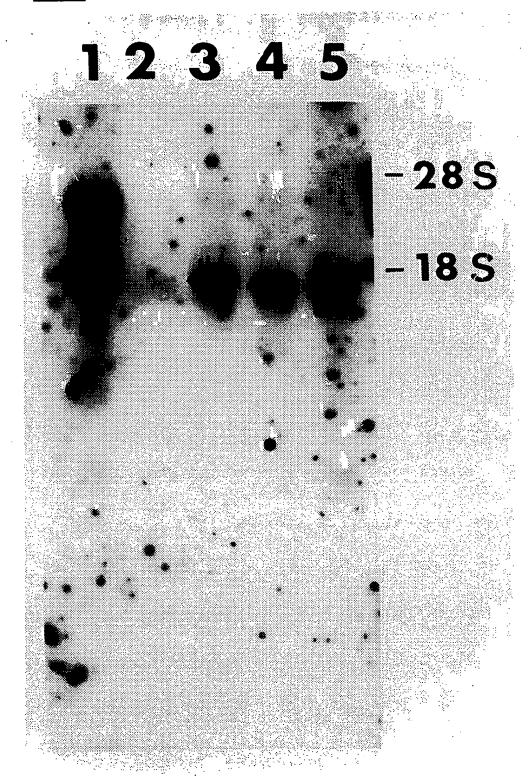
<u>Ļ</u>

ĸ

Sbjct: 501

Sbjct:

FIG 7



SUBSTITUTE SHEET (RULE 26)

FIG 8 (I) FIG 8 (II) FIG 8 (III) FIG 8 (IV) FIG 8 (V) FIG 8 (VI) FIG 8 (VII)

FIG 8

FIGURE 8 (I)

gb | AA074703 | AA074703 zm76g07.rl Stratagene neuroepithelium (#937231) 'n 531612 cDNA clone Homo sapiens 417 Length

Plus/Plus 206/259 (79%), Positives = 206/259 (79%), Strand = = 6.1e-103, Sum P(5)=6.1e-103Expect 818 (226.0 bits), Plus Strand HSPs: Identities = Score =

14/85

ţ

505 446 GGCCTCCCTCTGATGAGGTGGTGAGCCCCAGAGCCCCGAGCCCCTCAACACGTCTGAC Query:

GGGCTCCCTCTGATCGATGAGGTGATAAGCCCCAGAGCCCCGAGCCCCTCAATTCCTCAGAC 108

49

Sbjct:

565 506 TTCTCTGACTGGTCTAGTTTTAATGCCAGCAGTACCCCTGGACCAGAGGAGGTAGACAGC Query: 168 TTCTCTGATTGGTCCAGCTTTAATGCCACCACCACCTCTGTGCAAGAGGAGAGAGCCAGC 109

FIGURE 8(II)

Plus/Plus

398 2 Query: Sbjct:

œ

FIGURE

GCACTGAGAGAAAAGCTAGCCACAGTCAACTTGGCCCGGGCAGGACTGGGCTCCC

56

452

Sum P(5) = 6.1e-103Strand = 39/44 (88%), Expect = 6.1e-103, 11 39/44 (88%), Positives 175(48.4 bits),

810 GCCTTGGGTTGGCTGGCCCGGCTGCTAAGGAGCCGGGCTGGGTC 167 Query:

GCTCTGGGCTGGCCCAGCTGCTCAGGAGCCGGGCTGGGTC

416

Sum P(5) = 6.1e - 103(38.4 bits), Expect = 6.1e-103, 139 Score = Plus/Plus H Strand (888) 31/35 II Positives (888) 31/35 11 Identities

Substitute Sheet (Rule 26)

Identities

Score =

FIGURE

765 GGAGACTGTGACGATGACAAGTACCGACGTCGGCC Query: 731

GGAGACTGTGATGACAAATACCGCCGCCGGCC 336 Sbjct:

= Plus/Plus (36.8 bits), Expect = 6.1e-103, Sum P(5)=5.1e-103= 29/32 (90%), Positives = 29/32 (90%), Strand Score = 133

Identities

732

CGGGATGATGACCGGACAGCAGGCATTCATGG Sbjct: 305

AA134788 zm81g02.rl Stratagene neuroepithelium (#937231) 532082 gb | AA134788

FIGURE

Homo sapiens cDNA clone

368 Length

Strand HSPs: Plus (155.6 bits), Expect = 3.8e-87, Sum P(3)=3.8e-87563 Score = Plus/Plus 11 Strand (778), 147/190 Positives = 147/190 (778), 11 Identities

557 CGTCTGACTTCTCTGACTGGTCTAGTTTTAATGCCAGCAGTACCCCTGGACCAGAGGAGG 498 Query: CCTCAGACTTCTCTGATTGGTCCAGCTTTAATGCCACCACCACCTCTGTGCAAGAGGAGA 162 103 Sbjct:

FIGURE 8 (VI

617	222	677	282
ONATION 558 PAGACAGCCCCTCTGCCCCAGCCTTCTACAGCCAGGCCCCCGGCCCCCAGCTTCCC		CAGGCCGGCCCGAGCACACAGTGATCCACATGGGCAATCCTGAGCCCTTGACTCACG	
7,7,2	163	618	223
Onerv.	Sbjct:	Ouery:	Sbjct:

19/85

(95%), Strand = Plus/Plus Sum P(3) = 3.8e-873.8e-87, 94/98 Positives 11 Expect (125.4 bits), (828) 94/98 11 454 Identities Score =

CCCCAAGGAA 292

Sbjct: 283

678

Query:

	Query:	398	Query: 398 GCACTGAGAGAGAAGCTGGCCACAGTCAACTGGGCCCGGGCAGGACTGGGCCTCCCTC
S	Sbjct:	7	
ubstitute Sh	Query:		458 ATCGATGAGGTGAGCCCAGAGCCCCGAGCCCCTCAA 495
eet (Rule 2	Sbjct:	62	
6)	Score = 21 Identities	= 21 ities	= 219 (60.5 bits), Expect = 3.8e-87, Sum P(3)= 3.8e-87 ties = 51/60 (85%), Positives = 51/60 (85%), Strand = Plus/Plus
	Query:	702 (Query:702 GGGATGATGACCGGACACCAGGCCTCCATGGAGACTGTGACGATGACAAGTACCGACGTC 761

FIGURE 9

W32939 human

mouse AA242159

CTTCCGCGCGCTTTTCATTACCGTACGCACCGGTCA-CGATCGGCATCGCGGAGGATCGGTCATGGGACTTTGCAAG

FIG 10 (I)

FIG 10 (II)

FIG 10 (III)

FIG 10 (IV)

FIG 10

	•		
E) MGLCKCPKRK VTNLFCFEHR VNVCEHCLVA NHAKCIVQSY LQWLQDSDYN PNCRLCNIPL 60 ASRETTRLVC YDLFHWACLN ERAAQLPRNT APAGYQCPSC NGPIFPPTNL AGPVASALRE 120	* * * 	180 * DSASAAPAFY 60 a*tps****> 60 a*tps****> 180	60 sqh*icac*l>
I) MGLCKCPKRK VTNLFCFEHR VNVCEHCLVA NHAKCIVQSY LQWLQDSDYN PNCRLCNIPL ASRETTRLVC YDLFHWACLN ERAAQLPRNT APAGYQCPSC NGPIFPPTNL AGPVASALRE		170 * ASSTPGPEEV 50 *tt*svq**r 50 *tt*svq**r	50 chhhlcarge
NHAKCIVQSY APAGYQCPSC		160 * SDFSDWSSFN 40 ***********************************	40 s xrll*lvgl*
VNVCEHCLVA ERAAQLPRNT		150 * VSPEPEPLNT 30 * i******* 30 ags*s*sip	30 ** ** ** ** **
VTNLFCFEHR YDLFHWACLN		140 * GLGLPLIDEV 20 ******	20 ** * * —
10(I) MGLCKCPKRK ASRETTRLVC		130 * KLATVNWARA	10
FIGURE 10 MCG4 MCG4 3.	[229] 5. [74]	M CG4 1	3.

FIGURE 10 (II

24/85

	•		24705									
ω —	.60 ep*lhlxlli>	. 240	DDDKYRRRPA	120	*srhswetvm mtnt-aagl*>		•	120	^****	120	^U******	
	50 mppp*lckrr	230	рктрсінсьс	110				110	*******	110	kesi*h*gmm tgqqafm***	
	40 tslig-pal-	220	PRKVYDTRDD **	100	dbd******	^ * *	·rl —	100	*kek*m*hg*	100	kesi*h*gmm	ч
	30 q**s*-sipq	210	зкреднтиін мсмрерітна	06	**st*a*a** 90	**st*a*a**		06	gey*s*g*r-	06	gey*s*g*rp	·
<u> </u>	 ****smr**a	* 200	GRРЕОНТVІН	80	* * * * * * * * * * * * * * * * * * *	* * * * * * * * * * * * * * * * * * *		80	s*a-a*sht*	80	s*a-a*sht*	
	10 ******	190	SRAPRPPASP	70	* d * * * * p *	**Q*****D*		70	gsp^*sslpk^*	70	p*sslpk*	
	5.		MCG4		[372]	[243]		M	[229]	4.	[86]	

300 LGWLARLLRS RAGSRKRPLT LLQRAGLLLL LGLLGFLALL ALMSRLGRAA ADSDPNLDPL 290 2.80 <********** *4dgk*m*** 100 100 р * \ 270 arl*allppq av*sstqsyt w*vlk*w-*t 90 260 80 310 MNPHIRVGPS 250 ****D**** ×++-*0 130 FIGURE 10 (III) 98 38 MCG4 MCG4 9 Ŋ Substitute Sheet (Rule 26)

FIGURE 10 (IV)

Search Analysis for Sequence: MCG4 Search from 1 to 310

Date: September 22,1997

Maximum possible score: 1598

Score Region from 1 to 310

Matrix: pam250 matrix

Aligned sequences:

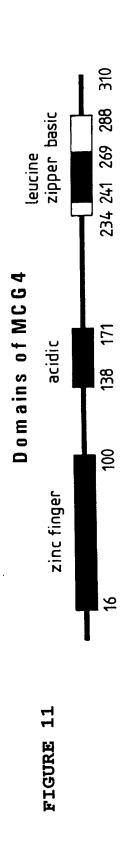
. = EST AA074703 phase 1 translation

EST AA134788 phase 3 translation EST AA134788 phase 2 translation

EST AA074703 phase 3 translation EST AA074703 phase 2 translation

5. = EST AAU/4/U3 phase 2 translation 5. = EST AA134788 phase 1 translation

II



consensus: CX2HX4CX2CX4HX2CX17CX2CX18HX2CX18CX2C zinc finger

0/34 acids, 9/34 negatively charged amino acidic domain consensus: positively charged

27/85

0/55

acids, amino positively charged 13/55 domain consensus: negatively charged basic

leucine zipper domain consensus: LX,LX,RX,LX,L

alternate "novel" leucine zipper-like motif where leucine would not be 261) (aa alpha helix domain: an of surface one LX₆LXLX₆LXLX₆L (aa 286) along the aligned

FIG 12 (I)	FIG 12 (II)
FIG 12 (III)	FIG 12 (IV)

FIG 12

FIG 12 (I)
Sequences producing High-scoring Segment Pairs:

(Z70752 F25B3.3 [Caenorhabditis ele (U53884) aimless RasGEF [Dictyosteli (U67326) Ras-GRF2 [Mus musculus] CDC25 protein homolog - yeast (Cand	CELL DIVISION CONTROL GUANINE NUCLEOTIDE RE	de-exhange-factor ho	mouse/gi 5	SCD25 PROTEIN /gi 457494	STE6 PROTEIN /pir 828098 stee CDC25 protein homolog - mouse	GUANINE NUCLEOTIDE RELEASING P	(s62035) Ras-specific guanine nucleo	SCD25 protein - yeast (Sacchar	CDC25 [Homo sapiens]	T14G10	(X03579) CDC25 protein (aa 1-1588) [
gn1 PID e236178 gi 1293099 gi 1655941	ol P43069 C	rr 1814463A r B46199	a1 PID e r s2269	p P14771 SC25	p P26674 STE i r s28407	p P27671	i 386047	p Q0234 ir s141	1143372	ni PID	i 3484

=1G 12 (II)

```
Sum

Sum

Probability

P(N)

N

3.0e-124

7.8e-22

5.2e-15

3.6e-16

4.6e-15

3.0e-14

3.0e-14

3.7e-14

3.7e-14

3.7e-14

3.0e-14

3.7e-14

3.7e-13

3.7e-13
```

High Score

1123 1153 1153 1153 1153 1153

VISION CONTROL PRO	ilorvegicus r homolog c	Molecule - excinange ractor momorog c	(U24071) Munc13-2 [Rattus norvegicus]		(U75361) Munc13-3 [Rattus norvegicus]	exchange fact	GUANINE NUCLEOTIDE DISSOCIATION STIM	LTE1 protein - yeast (Saccharomyces	(D21354) a putative guanine nucleoti	ERATURE E	$\overline{}$	꿏	PROTEIN KINASE C, BRAIN ISOZYME (PKC	protein kinase C (EC 2.7.1) delta	PROTEIN KINASE C, DELTA TYPE (NPKC-D	protein kinase C mu - human /pir A5	KINASE C, DELTA TYPE (NPK	(Z34524) serine/threonine protein ki	(U68142) RalGDS-like [Homo sapiens]
A la	i 91532	1	LC	1 47498	i 17633	i 80695	sp 00338	pir (BVB)	1145224	sp P0786	qi 509050	gi 52058	sp P0513	pir 535704	pl 0056	ir S40279	p P0921	1 52087	1 15197

=1G 12 (III

=1G 12 IV

```
7.2e-13
3.4e-12
3.4e-12
5.5e-12
1.5e-12
1.5e-11
1.0e-11
2.7e-11
2.7e-10
4.7e-10
4.7e-10
4.7e-10
4.7e-10
1.8e-09
3.8e-09
```

FIG 13a (I)

FIG 13a (II)

FIG 13a(III)

FIG 13a (IV)

FIG 13a (V)

FIG 13a (VI)

FIG 13a (VII)

FIG 13a (VIII)

FIG 13a (IX)

FIG 13a (X)

FIG 13a

44	89	134	179	224	269
FIGURE 13(a) (I) CG ATT TCA TTC CTC GCT CCC CAC AGG TCC CTC TCC CCA AAA TAT Ile Ser Phe Leu Ala Pro His Arg Ser Leu Ser Pro Lys Tyr 1 5	TCC CAT CTT GTC CTA GCC CAT CCC CCA GAC TAT CTC AAG GAC CAG Ser His Leu Val Leu Ala His Pro Pro Asp Tyr Leu Lys Asp Gln 15	CTG TCC CCA CGC CCC CGA CCT CCA CTA GGC CTG TGC CAC CCG CTG Leu Ser Pro Arg Pro Pro Leu Gly Leu Cys His Pro Leu 30	CCT GCA GGA AGA CGC CCG GTC CCG GGC CGG GTT AGC CCC ATG GGA Reproved Provel Provented Service Gly Arg Val Ser Provet Gly Service Gly Afs 45	ACG CAG CGC CTG TGT GGC CGC GGG ACT CAA GGC TGG CCT GGC TCA Thr Gln Arg Leu Cys Gly Arg Gly Thr Gln Gly Trp Pro Gly Ser 60	AGT GAA CAG CAC GTC CAG GAG GCG ACC TCG TCC GCG GGT TTG CAT Ser Glu Gln His Val Gln Glu Ala Thr Ser Ser Ala Gly Leu His 75

314	3 59		404	449	494	539	
AGG	Arg	GCC Ala	GAG Glu	GTG Val	TGG Trp	$ extsf{TAC}$	
ggg	Gly	GCG Ala	GAG Glu	AAG Lys	CCC	ATC Ile	
GGT	Gly	CCG Pro	GTG Val	$_{\rm GGG}$	CAC His	CAC His	
	Pro	GCG Ala	ACG Thr	TCC	ATG Met	CTC	
5 A 5	G1u 100	CCC Pro 115	TGC Cys 130	GAC ASP 145	ATG Met 160	CTG Leu 175	
ر E	Ser	CAC His	GGC G1y	GAT Asp	CTC	AAG Lys	
ָל ל	Arg	GCC Ala	AAG Lys	TTC Phe	TTC	GCC	
FF.	Val	CCA	GAC Asp	GCC Ala	ATG Met	GCG Ala	
ָר כ	$_{ m G1y}$	GGC G1y	CTG	GAA	CGC Arg	CTG Leu	
ָ ב	Leu 95	CTG Leu 110	GAC ASP 125	ATC Ile 140	GTG Val 155	CAG Gln 170	
<u> </u>	GAG Glu	AGC Ser	CTG	TGC	CTG Leu	TCT Ser	
H	GAC	CGC	ACC Thr	$_{\rm GGG}$	CAG Gln	TCC Ser	
13 (a)	GTG Val	GAG Glu	GGC Gly	CGC Arg	CCG	CCC Pro	
RE 1	$ ext{GGG}$	CCG Pro	GCA Ala	CTC	GAC Asp	ATC Ile	
FIGU	TCT GGG GTG GAC GAG C Ser Gly Val Asp Glu I 90	CTC Leu 105	ATG Met 120	CTG Leu 135	CGG Arg 150	TAC TYT 165	
			Substitute S	heet (Rule 26)			

584	1)	629	674	719	764	808
. کا م	Thr	TTT Phe	GCT	GAC Asp	CAG Gln	TTT Phe
AAA	Lys	GAG Glu	AAG Lys	ATC Ile	ACT Thr	TTG Leu
		GCG Ala	CTG	CTA	GTG Val	CTG Leu
ر م	Gln Val	CCA GCG Pro Ala	GAG	AGC Ser	CAG Gln	TCC Ser
	Leu 190	TTC Phe 205	AAG Lys 220	AGC Ser 235	CGG Arg 250	ATG Met 265
J L	Ser	GCC Ala	ATC Ile	CAC His	AAG Lys	AAG Lys
TAA	Asn	TCC	CAG	CGG Arg	TGG Trp	CGC Arg
ر H	Ser	ATC Ile	GAG Glu	CGA Arg	AAG Lys	AAG Lys
	Asn	TGG Trp	GCT Ala	AAC Asn	TAC	AAA Lys
ر لام	Asp 185	TAC Tyr 200	TTG Leu 215	GGG G1y 230	ACC Thr 245	CAG Gln 260
(HHH)	Lys	AGG Arg	GAG Glu	GAA Glu	CCT Pro	GGA G1y
H 77	Arg	GTC Val	CCG Pro	CAA Gln	GTC Val	GTG Val
[3 (a) ∏	Ser	CTG	AAC Asn	GAC Asp	AGC	CCT Pro
元氏 2 4 4 7 4 7 4 4 7 4 4 4 4 4 4 4 4 4 4 4	Gln	CAC His	TTG Leu	CTA Leu	GAC Asp	AAC Asn
FIGU	Gln Gln Ser Arg	TGC Cys 195	GAC ASP 210	CTG Leu 225	ATA Ile 240	CGG Arg 255
			Substitute Shee	(Rule 26)		

854	899	944	686	1034
TTG Leu	AGT Ser	CGG Arg	ATG Met	ACA Thr
TAC Tyr	CAC His	GAG Glu	CTC Leu	ATC Ile
ACC Thr	\mathtt{TAT}	CTG Leu	CAG Gln	GTC Val
CTC	GAC Asp	GTC Val	GTG Val	CTG
CAT His 280	CAG Gln 295	CCC Pro 310	TGG Trp 325	GCC Ala 340
GAG Glu	TTT Phe	GAC AAC Asp Asn	CAG Gln	CGG Arg
GCG Ala	CTG	GAC Asp	TCA	CAG Gln
CTG Leu	ATC Ile	GTG Val	GTC Val	CCG
GAG Glu	AAG Lys	ACT Thr	AAC AGC Asn Ser 320	GCC
ATG Met 275	TGC Cys 290	TGC Cys 305		ACA Thr 335
	TTC Phe	GGC Gly	TTC Phe	CCC Pro
GAG GAG Glu	TCC	CAT	CTC	AAA Lys
FIGURE 13(a) (IV) GAC CAC CTG GAG CCC ASP His Leu Glu Pro 270	CGC Arg	ACT Thr	TCC Ser	AGC
JRE 1 CAC His	TAT TYr	GTG Val	ATC Ile	CTC
FIGU GAC ASP 270	GAG Glu 285	TTC Phe 300	TTC Phe 315	ATC Ile

	1079		1124	1169	1214	1259	1304
	TTC	rne	ATC Ile	ATC Ile	AAC Asn	CGC Arg	CAG Gln
	AAC	ASII	TCC	ACC Thr	GGC Gly	TTC	CTG Leu
	CAG	6 L II	AGC Ser	GAG Glu	ACA Thr	GGC Gly	GCC Ala
	CTG	гeп	CAC His	CCT	GCG Ala	GTG Val	GTG Val
	CAG	355 355	AGC Ser 370	AGC Ser 385	ACG Thr 400	TGT Cys 415	CTG Leu 430
	CTA	ren	CTG	GTT Val	GTG Val	GCC	GAC Asp
	CTG	пел	GGC Gly	CAC	CTA Leu	GCA Ala	AAG Lys
	AAG	S □	666 G1y	AGC Ser	GAA Glu	CTG	CTC
	GAG	n T	GTC Val	CAC	ACG Thr	CGG	CAC His
	GCG 5 L &	A1a 350	GTG Val 365	ACC Thr 380	CTC Leu 395	CGT Arg 410	GTG Val
	GTG	Val	GCA Ala	GAG Glu	GGT Gly	CGG Arg	GGT G1y
S	CAC	HIS	ATG Met	AAG Lys	GAG Glu	TAC TY <i>r</i>	CTG
L3 (a)	GTC	Val	CTG	CTC Leu	TGG Trp	AAC Asn	ATC Ile
JRE	CAC TTT GTC	Phe	ACG Thr	CGC	CTC	GGC Gly	CCG Pro
FIG	CAC	H18 345	AAC Asn 360	TCC Ser 375	AAG Lys 390	TAT TYr 405	TTC Phe 420

1349	GCC 1394 Ala	CTG 1439 Leu	1484	TCC 1529 Ser
AAC Asn	GCC Ala	CTG	GAT Asp	TCC Ser
CTC	CTG	GAC Asp	GAG Glu	TCC AAG Ser Lys
CGG Arg	GAG Glu	CCC Pro	ACG Thr	TCC
ACC Thr	GAG Glu	AAC Asn	CAG Gln	CCG CGC Pro Arg 505
CCA GCC CGG Pro Ala Arg 445	CTG Leu 460	CCA CCA GTA CAG GCC AAC Pro Pro Val Gln Ala Ash 475	CTG GAT CAG TAT CAG Leu Asp Gln Tyr Gln 490	CCG Pro 505
GCC Ala	ATC Ile	CAG Gln	CAG Gln	GAG Glu
CCA Pro	CTC TTT AGC ATC Leu Phe Ser Ile	GTA Val	GAT ASP	CGG
GAC (TTT Phe	CCA	CTG Leu	CAG Gln
CTG	CTC Leu		TCT Ser	CTG
TGG Trp 440	CAG Gln 455	CGG Arg 470	GTG Val 485	TCC Ser
CCT GAC	G ATG AAG Cs Met Lys G	AGC CTG Ser Leu	CTC ACG Leu Thr	CTG Leu
CCT Pro	ATG Met			CAG
l 3 (a) CTG Leu	GCC AAG Ala Lys	ACC Thr	CTG	TAC
JRE 1 GCA Ala	GCC Ala	GTG Val	AGC Ser	CTG
FIGURE 13(a) CTG GCA CTG C Leu Ala Leu F	GGG G1Y 450	ATG Met	CTG Leu 480	GAG Glu
		Substitute Sheet	(Rule 26)	

	1574	1619	1664	1709	1754	1799
	CCG Pro	CAG	TTC Phe	GAA Glu	GGG G1γ	ATG Met
	CCC Pro	GAT Asp	GTG Val	GAA Glu	TTT Phe	GAG Glu
	CGG Arg	CTG Leu	TCT	CAG Gln	GCC Ala	GAG Glu
	CCC	AAG Lys	GAG Glu	TCA Ser	AGC Ser	AGG Arg
	CCA Pro 520	CCC Pro 535	GTG Val 550	ATC Ile 565	CTC Leu 580	AGC Ser 595
	CCA Pro	AAA Lys	ATG Met	CAC His	TAC Tyr	ATC Ile
	ACC Thr	GCC Ala	AAG Lys	GGC Gly	CCT	TGC
	TGC Cys	GCT Ala	GAG Glu	GAT Asp	TTC Phe	GGC Gly
	AGT Ser	TCG Ser	ATC Ile	$_{\rm GIY}^{\rm GGG}$	AAC Asn	GAT Asp
	ACG Thr 515	ACC Thr 530	CAC His 545	GAT Asp 560	GGG G1y 575	CAG Gln 590
Ĥ	CCC	TGG Trp	GAG Glu	GTC Val	CGT	AAC Asn
IA)	AGC Ser	GAG Glu	GTG Val	GAC Asp	ATC Ile	CAG Gln
13(a)	ACC	GAG Glu	GTG Val	TTT Phe	ATC Ile	GAC
	4 0	CTG	CTC	AAC Asn	CAG Gln	CTC
FIGURE	TCG Ser 510	GTA Val	GCC Ala 540	CGG Arg 555	TTC Phe 570	GAC Asp 585
			Carboniana Cha	et (Rule 26)		

Substitute Sheet (Rule 26)

1844	1889	1934	1979	2024
ATG . Met	GTC Val	CAG Gln	TGC Cys	GTG Val
CGC Arg	CCC	AAG Lys	CAG Gln	AGT
666 G1y	CGC Arg	TAC TYr	AAG Lys	GCC CAG Ala Gln
GGG G1γ	TTG	ATC Ile	CAC His	GCC Ala
TTG Leu 610	TCC Ser 625	GGC G1y 640	TGC Cys 655	AGG Arg 670
GTG Val	AAC Asn	CTG Leu	AAC Asn	CGC
TCT Ser	AGC Ser	ATC Ile	GGA GTG Gly Val	CGG
AGC Ser	GAG Glu	CTG		TGT
TCC Ser	CAG Gln	GCC Ala	TGT Cys	GAG Glu
CGC Arg 605	TTC Phe 620	AAA Lys 635	GCC Ala 650	GTT Val 665
CTG CTG Leu	AAC Asn	TGC	CGA Arg	TCA
TTC Phe	CAC His	CAC His	TGC	CTG
FIGURE 13(a) (VILL) GTT TCC TAT TTC CTG Val Ser Tyr Phe Leu 600	GTA Val	CGC Arg	CTC AAA TGC CGA Leu Lys Cys Arg	CGC
JAB 1 TCC Ser	TTC	TGC Cys		GAT ASP
Fig GTT Val 600	GGC G1Y 615	GCC Ala 630	GGC G1Y 645	AAG Lys 660

CAC 2069		CGA 2114 Arg	GTG 2159 Val	2198	2248	2293	0770
FIGURE 13(a) (IX) AGC CTG GAG GGG TCT GCA CCC TCA CCC ATG CAC AGC CAC	Ser Ala Pro Ser Pro Ser 680	CAT CAC CGC GCC TTC AGC TTC TCT CTG CCC CGC CCT GGC AGG His His Arg Ala Phe Ser Phe Ser Leu Pro Arg Pro Gly Arg 690	GGC TCC AGG CCT CCA GAG ATC CGT GAG GAG GTA CAG ACG Gly Ser Arg Pro Pro Glu Ile Arg Glu Glu Glu Val Gln Thr 705	GAG GAT GGG GTG TTT GAC ATC CAC TTG TA ATAGATGCTG Glu Asp Gly Val Phe Asp Ile His Leu * 720	TGGTTGGATC AAGGACTCAT TCCTGCCTTG GAGAAAATAC TTCAACCAGA	GCAGGGAGCC TGGGGGTGTC GGGGCAGGAG GCTGGGGATG GGGGTGGGAT	この作を担びりの目が、目がり目がりのできた。 ひかかり きのかか きのは きがっぱい 田がり きゅうしょう
			Substitute She	et (Rule 26)			

2398	2416
ATAAAAAGGC	
TCCAGATGGA	
ATTTGTATTT	n(
rgergaat	AACCTTC (A
FIGURE 13(a) TIGTACAGAC TC1	CCGTGTAATT

FIGURE 13(b)		
CGATTTCATT CCTCGCTCCC CACAGGTC	CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT	20
CTTGTCCTAG CCCATCCCCC AGACTATC	CTTGTCCTAG CCCATCCCCC AGACTATCTC AAGGACCAGC TGTCCCCACG	100
CCCCCGACCT CCACTAGGCC TGTGCCACCC	SCC TGTGCCACCC GCTGCCTGCA GGAAGACGCC	150
CGGTCCCGGG CCGGGTTAG CCC CAT G	CCGGGTTAG CCC CAT GGG AAC GGG GTT CGG TCC GAG	196
* Pro His G	Pro His Gly Asn Gly Val Arg Ser Glu	
	1 5	
CCC GGT GGG AGG CTC CCG GAG CG	CCC GGT GGG AGG CTC CCG GAG CGC AGC CTG GGC CCA GCC CAC	238
Pro Gly Gly Arg Leu Pro Glu Ar		
10 15	15 20	
CCC GCG CCG GCG GCC ATG GCA GG	S ATG GCA GGC ACC CTG GAC AAG	280
Pro Ala Pro Ala Ala Met Ala Gl	a Met Ala Gly Thr Leu Asp Leu Asp Lys	
25 30	30 35	-
GGC TGC ACG GTG GAG GAG CT		300
Gly Cys Thr Val Glu Glu Leu	Glu	
40		

PCT/AU98/00380

45/85

FIG 14 (I)

FIG 14 (II)

FIG 14

MAGTLDLDK . MSSV17FFDO
SSKVEEDQHQELLIEDQLVARCVECF <u>UVDGAGVGAGT</u> VDHUFUSIIQW
IPSSQLAAKLLHIYQQS .
LSDSLSLITHFVNFYQETRNVEQREAVCRAVSFWIEKFPMHFDAQPQ
LAEQIKELKALLDQEGNRRHSSLIDIDSVPTYKWKRQVTQRNPVGQKK
CAQVVRLKTIAEDINE
RKMSLLFDHLEPMELAEHLTYLEYR
RVDFETLPTPGTPPFFPIASKKFSLTAFSLSFVQASPSDISTSLSHIDYR
SFCKILFQDYHSFVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTAP
LSTISITELKQYVKDGF
ORALVITHFVHVAEKLLOLONFNTLMAVVGGLSHSSISRLKETHSHVSPE
ERAEILV <u>KFVHVAKHLRKINNFNTLMSVVGGITHSSVARLAKTY</u> AVLSND
TIKLWEGLTELVTATGNYGNYRRRLAAC.VGFRFPILGVHLKDLVALQLA
KKELTQLTNLLSAQHNF

Substitute Sheet (Rule 26)

318	LPDWLDPARTRINGAKMKQLFSILEELAMVTSLRPPV.QANPDLLSLLTV	366
	GANFEKT. KCISSDKLVKLSKLLSNFLVFNQKGHNLPEMNMDLINTLKV	394
	SLDOYOTEDELYQLSLOREPRSKSSPTSPTSCTPPPRPPVLEEWTSAAKP	416
	. : : : SLDIRYNDDDIYELSLRREPKTFMNFEPSRGLVFAEWASGVTV	437
	KLDQALVVEHIEKMVESVFRNF <u>DVDGDGHISQEEF</u> QIJRGNFPYLSAFGD	466
	APDNATVSKHISAMVDAVFKHY <u>DHDRDGFISOEEF</u> QLIAGNFPFIDAFVN	487
	LDONODGCI SREEMVSYFLRSS. SVLGGRMGFVHNFOESNSLRPVACRHC	515
	: : : : : . : : :	537
	KALTLGTYKOGLKCRACGVNCHKOCKDRLSVECRRRAQSVSLEGSAPSPS	565
	.	587
	NALLEWGI DOOG DOOD OF THE WASHINGTON OF THE WASH	
	PMHSHHHRAFSFSLPRPGRRGSRPPEIREEEVQTVEDGVFULHL	60 <i>8</i>
	PRGSMRSKIINTCNNSGSTPDEEIGLVSLACEEVFEDDDL	627

Substitute Sheet (Rule 26)

FIG 15 (I)

FIG 15 (II)

FIG 15

FIGUR	FIGURE 15 (I)						
human	human CGATTTCATT CCTCGCTCCC CACAGGTCCC TCTCCCCAAA ATATTCCCAT CTTGTCCTAG	CCTCGCTCCC	CACAGGTCCC	TCTCCCCAAA	ATATTCCCAT	CTTGTCCTAG 60	
 human		CCCATCCCCC AGACTATCTC	AAGGACCAGC	AAGGACCAGC TGTCCCCACG CCCCCGACCT		CCACTAGGCC 120	
 human	TGTGCCACCC	GCTGCCTGCA	TGTGCCACCC GCTGCCTGCA GGAAGACGCC CGGTCCCGGG		CCGGGTTAGC	CCCATGGGAA 180	
human	CGCAGCGCCT	GTGTGGCCGC	GGGACTCAAG	GCTGGCCTGG	CTCAAGTGAA	GIGIGGCCGC GGGACTCAAG GCIGGCCTGG CTCAAGIGAA CAGCACGICC 240	
suom se			***tcag**	****ag****	L******	***a*g***t>	
 insering in the second in the	AGGAGGCGAC	CTCGTCCGCG	GGTTTGCATT	CTGGGGTGGA	CGAGCTGGGG	GTTCGGTCCG 300	
JTE SHEET (R					acagg ————	JTE SHEET (R	
esnor	g****t**a	**-*catt**	*****	***aa**aa*	g**ct****	**a**aat**>	
 human	human AGCCCGGTGG GAGGCTCCCG GAGCGCAGCC	GAGGCTCCCG	GAGCGCAGCC	TGGGCCCAGC	TGGGCCCAGC CCACCCGCG	CCGGCGGCC <u>A</u> 360	
mouse	****	*****tga	***t*t*a*t	****t*t**	***-*tg**a	^*******	
 human	human <u>IG</u> GCAGGCAC CCTGGACCTG GACAAGGGCT GCACGGTGGA GGAGCTGCTC	CCTGGACCTG	GACAAGGGCT	GCACGGTGGA	GGAGCTGCTC	CGCGGGTGCA 420	
 esnow	****ga****	L******	*******	******	*****	**t**C**t*>	
 human	1 TCGAAGCCTT	CGATGACTCC	GGGAAGGTGC	GGGACCCGCA	GCTGGTGCGC	GGGAAGGTGC GGGACCCGCA GCTGGTGCGC ATGTTCCTCA 480	

TOSTA	FIGURE IS (III)					
mouse	****** asnow	t******t	*******	*a**t**a**	*******	*****
human	TGATGCACCC	CTGGTACATC	CCCTCCTCTC	AGCTGGCGGC	CAAGCTGCTC	human IGAIGCACCC CIGGIACAIC CCCICCICIC AGCIGGCGGC CAAGCIGCIC CACAICIACC 540
mouse	*****	D*******	**+**	******	g**a****	***t***t*
human	human AACAATCCCG GAAGGACAAC TCCAATTCCC TGCAGGTGAA AACGTGCCAC	GAAGGACAAC	TCCAATTCCC	TGCAGGTGAA	AACGTGCCAC	CTGGTCAGGT 600
esnow	********	*****	*********	******	****	L******
ghuman	ACTGGATCTC	CGCCTTCCCA	GCGGAGTTTG	ACTTGAACCC	GGAGTTGGCT	g ghuman actggatete egeetteeea geggagtte acttgaaeee ggagtteget gageagatea 660
e monse	*****	3******	********	****	*******	<
ahuman Phuman	AGGAGCTGAA	GGCTCTGCTA	GACCAAGAAG	GGAACCGACG	GCACAGCAGC	$rac{\mathbb{R}}{2}$ human AGGAGCTGAA GGCTCTGCTA GACCAAGAAG GGAACCGACG GCACAGCAGC CTAATCGACA 720
e monse	*****	****	****	******	*****	^*********
human	TAGACAGCGT					730
mouse	*C**g**t**					

FIG 16 (I)

FIG 16 (II)

FIG 16 (III)

FIG 16

50	ω	140	182	224	266	308
FIGURE 16 (I) CACGCCTCGG AAGGGAGGTT TGGGGTCGGT GGTTTCACAG TGAGTGTGTC	CCGTTACCCG	GTG GGG ACC CCA ACC GCC TGC GGC TGC CCC TCC CAA GTT CCT Val Gly Thr Pro Thr Ala Cys Gly Cys Pro Ser Gln Val Pro 5	CCC TGT TGG CCA GGC ATC CAG GTC TCC AGT CTC CGA GCT GCG F Pro Cys Trp Pro Gly Ile Gln Val Ser Ser Leu Arg Ala Ala 20	GAG AAC CCA CCA CAT GCG GCT GCC CCT TTC CAT TCG ACC Glu Asn Pro Pro His Ala Ala Ala Pro Phe His Ser Thr	CTG TGG GGA GCC AGG CTT CCG GGG CCC CGT TCC TCC TGT GTG Leu Trp Gly Ala Arg Leu Pro Gly Pro Arg Ser Ser Cys Val 50	AAC TGG GCC CCC CCC CAT TCC CAG ACA TCA AGG CCG CGT Asn Trp Ala Pro Arg Pro His Ser Gln Thr Ser Arg Pro Arg 60

	350	392	434	476	518	
	· •					
	.	4.0	4 0 10	<i>r</i>)	()	> .
	TCC Ser	CCA Pro	CCA Pro 115	GTC Val	GGC	G1y
	AGG Arg	CCC Pro 100	CCT Pro	CCG Pro	\mathtt{TGT}	Cys
	CAC His 85	CAT	CGA Arg	CGC Arg	CTG	Leu
	CCC	GCC Ala	CCC Pro	AGA Arg	CGC	Arg 140
	GCT Ala	CTA Leu	CGC Arg	GGA G1Y 125	CAG	Gln
	CTC	GTC Val	CCA Pro 110	GCA Ala	ACG	Thr
	TTC Phe	CTT Leu 95	TCC Ser	CCT Pro	GGA G1v	G1y
	TCA Ser 80	CAT His	CTG Leu	CTG Leu	ATG	Met
	ATT Ile	TCC Ser	CAG Gln	CCG Pro	CCC	Pro 135
	ACG Thr	\mathtt{TAT}	GAC Asp	CAC His 120	AGC *	Ser
(H	<u>ರ</u> . ಹ	AAA Lys	AAG Lys 105	TGC Cys	${ m GTT}$	Val
I) 9	ATÀ [1e	CCA AAA Pro Lys 90	CTC	CTG	550	Arg
RE 1	CAG Gln 75	TCC	TAT Tyr	CTA GGC CTG Leu Gly Leu	ອອວ ວອອ ອວວ	$_{ m G1y}$
FIGU	CTC CAG ATA GC Leu Gln Ile Al 75	CTC	GAC Asp	CTA Leu	SCG	Pro Gly Arg 130
- '	_					

Substitute Sheet (Rule 26)

	260	602	644	989	720
	GTC Val	GAC Asp	GAG Glu 180	GCA Ala	
	CAC His	GTG Val 165	CCG	<u>ATG</u> Met	
	CAG Gln 155	GGG GTG Gly Val 165	CTC	GCC Ala	_U
	GAA Glu	TCT Ser	GGG AGG Gly Arg	GCG	GTG Val 205
	AGT Ser	CAT His	$^{\rm GGG}_{\rm G1Y}$	CCG Pro 190	ACG Thr
	GGC TCA AGT Gly Ser Ser	TTG	GGT G1Y 175	GCG Ala	TGC Cys
	GGC G1Y	GGT G1y 160	CCC Pro	CCC Pro	GGC Gly
	CCT Pro 150	GCG Ala	GAG Glu	CAC His	AAG Lys
	TGG Trp	TCC	TCC Ser	GCC	GAC ASP 200
	GGC G1y	TCG	CGG	CCA Pro 185	CTG
(HH	CAA Gln	ACC	GGG GTT Gly Val 170	GGC G1y	GAC
	CGC GGG ACT CAA Arg Gly Thr Gln 145	GCG Ala 155	GGG G1y	CTG	CTG
RE 1	GGG G1Y 145	GAG	CTG	AGC	ACC Thr
FIGU	CGC GGG Arg Gly 145	CAG Gln	GAG Glu	CGC Arg	GGC G1y 195

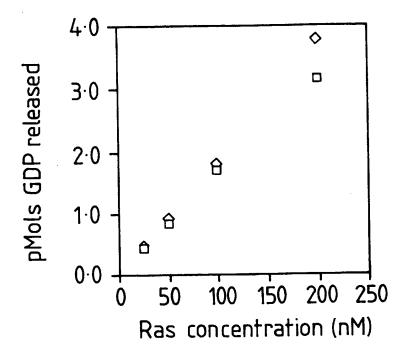
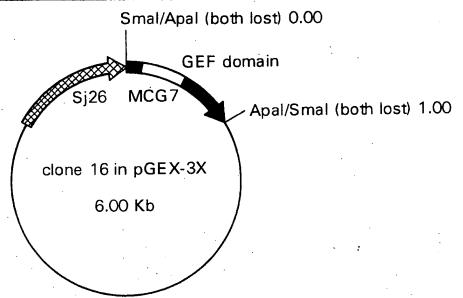


FIGURE 17

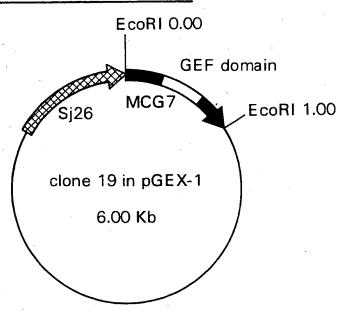
56/85 FIGURE 18 (Cont. I)



Plasmid name: clone 16 in pGEX-3X

Plasmid size:6.00 kb

FIGURE 18 (Cont. II)

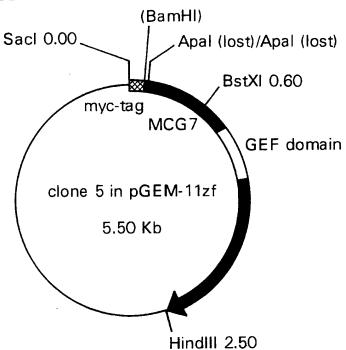


Plasmid name: clone 19 in pGEX-1

Plasmid size: 6.00 Kb

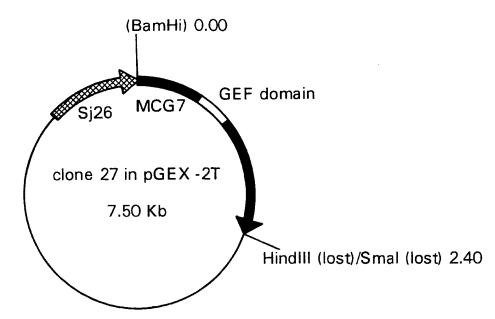
WO 98/53061 PCT/AU98/00380

57/85 FIGURE 18 (Cont. III)



Plasmid name: clone 5 in pGEM-11zf

Plasmid size: 5.50 kb



Plasmid name: clone 27 in pGEX-2T

Plasmid size: 7.50 kb

FIGURE 18 (Cont. IV)

SUBSTITUTE SHEET (RULE 26)

FIG 19 (I)

FIG 19 (II)

FIG 19 (III)

FIG 19 (IV)

FIG 19

43	8 2	127	169	211	253
FIGURE 19 (I) GCCCGCCGCC ATG CCG CCC TTA CTG CCC CTG CGC CTG TGC CGG Met Pro Pro Leu Pro Leu Arg Leu Cys Arg	CTG TGG CCC CGC AAC CCT CCC TCC CGG CTC CTC GGA GCG GCC Leu Trp Pro Arg Asn Pro Pro Ser Arg Leu Leu Gly Ala Ala 25	GCC GGG CAG CGG TCC AGA CCC AGT ACT TAT GAA CTG TTG Ala Gly Gln Arg Ser Arg Pro Ser Thr Tyr Tyr Glu Leu Leu 30	GGG GTG CAT CCT GGT GCC AGC ACT GAG GAA GTT AAA CGA GCT GGG GTG CAT CAT GTG GTG GAA GTT AAA CGA GCT GGG GAA GTT AAA CGA GCT GGG GTG CAT AAA CGA GCT GGG GAA GTT AAA CGA GCT GGG GAA GTT AAA CGA GCT GGG GTG GTG AAA CGA GCT AAA CGA GCT AAA CGA GCT AAA CGA GCT AAAA CGA GCT AAAAAAAAAA	TTC TTC TCC AAG TCC AAA GAG CTG CAC CCA GAC CGG GAC CCT Phe Phe Ser Lys Glu Leu His Pro Asp Arg Asp Pro 60	GGG AAC CCA AGC CTG CAC AGC CGC TTT GTG GAG CTG AGC GAG Gly Asn Pro Ser Leu His Ser Arg Phe Val Glu Leu Ser Glu 75

295	337	379	421	463	505
TAT Tyr 95	CGA	TCC	AGC Ser	AAA Lys	CTG Leu 165
AGC Ser	CCA	AGC Ser	CAC	CAC His 150	ATG Met
CGC Arg	TCT Ser	CAC	TTT Phe 135	CAA Gln	CTC
CGC Arg	AAG Lys	ACA Thr 120	CAG Gln	CAG Gln	CTC
AGC	CCA Pro 105	CAA Gln	TCC Ser	CAG Gln	CTC
CAG Gln 90	CCC Pro	CAC His	TGG	AGG Arg	TGC Cys
GAG Glu	AGT Ser	GCC Ala	\mathtt{TAC}	TTG Leu 145	TAC TYr
CGT	GGT Gly	TCT Ser	CAG Gln 130	CAG Gln	GGG G1y
AGC Ser	TCA Ser	AAG Lys 115	GCA Ala	CCC	CTG
CTC	CGC Arg 100	GAC Asp	AAC Asn	GGG G1γ	GTG Val
(rr) Tr GTG G Val	CTC	CAT His	CCC Pro	CAG Gln	CAA Gln 155
. 9 (1 CGT Arg	CAG Gln	GTC Val	CCC Pro	CCA Pro 140	AAA Lys
RE 1 TAC TYF	GAC Asp	ACA Thr	ACA Thr 125	AGG Arg	AAC Asn
FIGURE 19 (GCA TAC CGI Ala Tyr Arg	GAT Asp	ACC Thr 110	TGG	GTG Val	CAA Gln
		Substitute She	et (Rule 26)		

Substitute Sheet (Rule 26)

547	589	631	673	715	763
					Į.
AAG Lys	ATC Ile	AAC Asn	CGG Arg	GTG Val 235	GGCCTGCAGT
GTG Val	ATC Ile	GCC Ala	CAG Gln 220	ATC Ile	3600.
	CGG Arg	AGG Arg 205	GGG CAG Gly Gln 220	CCC GAG ATC Pro Glu Ile	
AGG AAG Arg Lys	GAT ASP 190	GCC Ala	CTA Leu	CCC	ACCTGGATGG
TTC Phe 175	AAG Lys	CGG Arg	CGG Arg	GGC (ACCI
GCC Ala	GAA Glu	GCA	CAA	ACC CAA (Thr Gln (230	CTC
ATT Ile	GAT Asp	CGG Arg	CGA Arg 215	ACC	TGA GGGGCTC *
\mathtt{TAC}	ATG Met	GCC Ala 200	GAG Glu	CCA	TGA *
CAC His	TTC Phe 185	GAA Glu	CAG Gln	GAG Glu	CCC Pro
CTG Leu 170	AAC Asn	AAC Asn	CAG Gln	TCC Ser	GGC G1y 240
111) GGC G1Y	CTT	\mathtt{TAC}	CTT	CCA Pro 225	GCC Ala
.9 (I ATG Met	CAC His	TTC Phe	ATC I1e 210	CCA	$ ext{GGC}$
JURE 19 (1 GGC ATG Gly Met	ATG Met	GCC Ala 195	GGC Gly	CCG Pro	CGG Arg
FIGURE 19 (III) GCG GGC ATG GGC Ala Gly Met Gly	CAG Gln 180	ACA Thr	AGA Arg	CAG Gln	CCC
		Substitute SI	heet (Rule 26)		

832

FIGURE 19 (IV) GCGTTCCCGC TTTGCTTCCT TCCCTGGACG GCCCGCTCCC CGAAACGCGC

GCAATAAAGT GATTCGCAG

FIGURE 20

shock protein >sp|P08622|DNAJ_ECOLI DNAJ PROTEIN >pir ||HHECDJ heat

dnaJ -

Escherichia coli >gi |145769 (M12565)-heat shock protein dnaJ

[Escherichia coli] >gi |216441 (D10483) dnaJ protein

[Escherichia coli]

Length = 376

= 1.2e-10Score = 138 (63.7 bits), Expect = 1.2e-10, P

63/85

39/62 (62%) 11 25/62 (40%), Positives 11 Identities Query: 35 YYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRS94 ++F E+ EAY VL+ + HPDR+ G+ E+++A+ Ø YYE+LGV

YYEILGVSKTAEEREIRKAYKRLAMKYHPDRNQGDKEAEAKFKEIKEAYEVLTDSQKRAA 65 ဖ Sbjct:

Query: 95 YD 96

YD

Sbjct: 66 YD 67

Substitute Sheet (Rule 26)

FIG 21 (I)

FIG 21 (II)

FIG 21 (III)

FIG 21 (IV)

FIG 21

FIGURE 21 (I

DNAJ-like domain ಡ to similarity contains (Caenorhabditis elegans] >gi|1703590 (U80439)

Length = 345

5.2e-12 11 5.2e-12, Sum P(3) 98 (45.2 bits), Expect = Score =

28/37 (75%) 11 17/37 (45%), Positives 11 Identities

64 QRSRPSTYYELLGVHPGASTEEVKRAFFSKSKELHPD Query: 28

E+K AF+++SK++HPD

A+

T+YE+LGV

++ R

58 KKIRQRTHYEVLGVESTATLSEIKSAFYAQSKKVHPD 22 Sbjct:

5.2e-12 II 74 (34.1 bits), Expect: = 5.2e-12, Sum P(3) 11 Score

= 17/32 (53%), Positives = 19/32 (59%)Identities

FIGURE 21 (II

102 SLHSRFVELSEAYRVLSREQSRRSYDDQLRSG Query: 71

S + F+EL AY VL R RR YD QLR G

95 SATASFLELKNAYDVLRRPADRRLYDYQLRGG 64 Sbjct:

II 5.2e-12, Sum P(3) (18.0 bits), Expect = 39 11 Score

= 10/42 (23%), Positives = 19/42 (45%)

203 LLMLAGMGLHYIAFRKVKQMHLNFMDEKDRIITAFYNEARAR O L+ χ+ L+++AG 199 LVLVAGYNGGYLYLLAYNQKQLDKLIDEDEIAKCFLRQKEFR 158 Sbjct:

Identities

162

Query:

elegans] >gnl |PID |e281266 (Z81030) COlG10.12 [Caenorhabditis Length = 191

FIGURE 21 (III)

1.8e-0911 96 (44.3 bits), Expect = 1.8e-09, Sum P(3) II Score

27/41 (65%) Identities = 17/41 (41%), Positives = 75 35 YYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSR Query:

SRK+K+LHPD+ A+ +E++ AF YYE++GV

59 YYEIIGVSASATRQEIRDAFLKKTKQLHPDQSRKSSKSDSR 19 Sbjct: 54 (24.9 bits), Expect = 1.8e-09, Sum P(3)= 10/22 (458), Positives = 15/22 (688)Identities II Score

11

96 Query: 75 RFVELSEAYRVLSREQSRRSYD

E+ R+ YD +F+ + EAY VL

(IV) FIGURE 92 **QFMLVKEAYDVLRNEEKRKEYD** Sbjct: Expect = 1.8e-09, Sum P(3) (16.1 bits), 35 II Score

22/44 (50%) 11 9/44 (20%), Positives 11 Identities QGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFRKVKQMHLN 184 + KQ + P+ 141

++A +G ++I RNPEDEYLREKOKNRMLVVLAATVMALIGANIVYIRKLQADRLS

188

Substitute Sheet (Rule 26)

Query:

145

Sbjct:

WO 98/53061 PCT/AU98/00380

69/85

FIG 22 (I)

FIG 22 (II)

FIG 22 (III)

FIG 22

FIGURE 22 (I)

ZI >sp|Q10209|YAY1_SCHPO HYPOTHETICAL 44.8 KD PROTEIN C4H3.01 CHROMOSOME I

pombe] >gi|1184014 (269380) unknown [Schizosaccharomyces

Length = 392

11 Sum P(3) (38.8 bits), Expect = 4.1e-08, 84 11 Score

(869) 25,36 11 Positives 13/36 (36%), H Identities 70 35 YYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNP Query:

YY+LLG+ A+ ++K+A+ + + HPD++P +P

44 YYDLLGISTDATAVDIKKAYRKLAVKYHPDKNPDDP Sbjct:

II = 4.1e-08, Sum P(3) (29.5 bits), Expect 64 11 Score

23/40 (57%) !! = 14/40 (358), Positives Identities

11

RFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRTTVHD 75 Query:

FIGURE 22 (II)

Q+ Н + P+ YD

114

E+ R ++SEAY+VL 十 円

89 KFQKI SEAYQVLGDEKLRSQYDQFGKEKAVPEQGFTDAYD 50 Sbjct:

Identities

15/29 (51%) II 9/29 (31%), Positives н

Score = 37 (17.1 bits), Expect = 4.1e-08, Sum P(3)

190 DRIITAFYNEARARARANRGILQQERQRL Query:

RQR+ +++ + A A+ 闰 Ø

DR

DRKKNAQIREREALAKREQEMIEDRRQRI 149 Sbjct: 0.00081 II 0.00081, Sum P(3) 11 (15.2 bits), Expect 33 H Score

8/19 (42%), Positives = 11/19 (57%) 11 Identities

FIGURE 22 (III)

) }		62
CAL CALKANTAXXIIIXXXIIIX 1001 041	QVLG	POGASEKFOKTSFAVOVIG 62
ָּבְּאָבְאָ בְּאָבָאָבָ	+	SEA.
יג גע	+ŏ +	OKT.
X 1 1 7	+	NEKE TKE
r XGEV	PQG	מ טטם
T T		7
Yuer y.	•	Chict. 11

FIGURE 23

tumorous imaginal discs [Drosophila virilis] (Y07700) Tid58 protein [Drosophila virilis] >gn1|PID|e263866 >gn1|PID|e253406 (X77635) 529 11 Length

9.7e-13 11 9.7e-13, P score = 153 (70.6 bits), Expect =

44/71 (618) 11 Positives 27/71 (38%), li Identities

73/85

Query: 26

85 AGQRSRPSTYYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRV ++SEAY

五十 +P + +K+ HPD A+ +++K+A++ ΓGV XX召

> 72 Sbjct:

+

131 SSSRMQAKDYYATLGVAKNANAKDIKKAYYELAKKYHPDTNKDDPDASKKFQDVSEAYEV

96 LSREQSRRSYD 98 Query:

LS +Q RR YD

LSDDQKRREYD 142 Sbjct:132

FIG 24 (I)

FIG 24 (II)

FIG 24 (III)

FIG 24

FIGURE 24(I)

75/85

MCG18 HDJ-2 HDJ-2 HDJ-2 HDJ-2 HDJ-1 HDJ-2	MVKETTYYDVLGVKPNATQEELKKAYRKLALKYHPDKNPPSRLLGAA MVKETTYYDVLGVKPNATQEELKKAYRKLALKYHPDKNPNEGEKFKQISQAYEV MGKDYYQTLGLARGASDEEIKRAYRRQALRYHPDKNNPNEGEKFKQISQAYEV M-ASYYEILDVPRSASADDIKKAYRRKALQWHPDKNPDNKEFAEKKFKEVAEAYEV AGQRSRPSTYYELLGVHPGAST-EEVKRAFFS- LSDAKKRELYDKGGEQAIKEGGAGGGFGSPMDIFDMFFGGG LSDAKKREIFDRYGEEGLKGSGPSGGSGGGANGTSFSYTFHGDPHAMFAEFFG LSDKHKREIYDRYGREGLTGTGTGFSRAEAGSGGPGFTFT-FRSPEEVFREFFG KSKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRT GRMQRERRGKNVVHQLSVTLEDLYNGATRKLALQKNVICDKCEGRGGKKGAVECCPNCRG GRNDFDTFFGQRNGEEGMDIDDPFSGFPMGMGGFTNVNFGRSRSAQEPARKKQDPPVT
HSJ1	SGDPFAELFDDLGPFSELQNRGSRHSGPFFTFSSSFPGHSDFSSSSFSFSPGAGAFRS

Substitute Sheet (Rule 26)

FIGURE 24 (II)

٠	MCG18	TVHDKSAHQTHSSWTPPNAQYWSQFHSVRPQGPQLRQQQHKQN
	HDJ-2	TGMQIRIHQIGPGMVQQIQSVCMECQGHGERISPK-DRCKSCNGRKIVREKKILEVHIDK
	HDJ-1	HDLRVSLEEIYSGCTKKMKISH-KRLNPDGKSIRNEDKILTIEVKK
S	HSJ-1	VSTSTTFVQGRRITTRRIMENGQ-ERVEVEEDGQLKSVTINGVPD *
ubstitu	MCG18	KQVLGYCLLLMLAGMGLHYIAFRKVKQMHLNFMDE-KDRIITAFYNEARARAN
te She	HDJ-2	GMKDGQKITFHGEGDQEPGLEPGDIIIVLDQKDHAVFTRRGEDLFMCMDIQLVEALCGFQ
et (Ri	HDJ-1	GWKEGTKITFPKEGDQTSNNIPADIVFVLKDKPHNIFKRDGSDVIYPARISLREALCGCT
de 26)	HSJ1	DLARGLELSR-REQQP-SVTSRSGGTQVQQTPASCPLD-SDLSEDEDLQLAMAYSLSE * *
	MCG18	RGILQQERQRLGQRQPP-PSEPTQGPEIVPRGAGP
	HDJ-2	KPISTLDNRTIVITSHPGQIVKHGDIKCVLNEGMPIYRRPYEKGRLIIEFKVNFPENGFL
	HDJ-1	VNVPTLDGRTIPVVFKDVIRPGMRRKVPGEGLPLPKTPEKRGDLIIEFEVIFPERI
	HSJ1	MEAAGKKPAGGREAQHR-RQGRPRPSTKIQAWGGPRRVRGVKQPNAVHPQR-RR

FIGURE 24 (III)

	SPDKLSLLEKLLPERKEVEETDEMDQVELVDFDPNQERRRHYNGEAYEDDEHHPRGGVQC	PQTSRTVLEQVLPI	PLAASSSEHRAQPDLIQILTGGSDSLWEEKRGVS
MCG18	HDJ-2	HDJ-1	HSJ1

* = amino acid identity in all 4 proteins

conservative substitution

II

Substitute Sheet (Rule 26)

HSJ1

 Ω TS

FIG 25 (I)

FIG 25 (II)

FIG 25(III)

FIG 25 (IV)

FIG 25

47	68	131	173	215	257
CCG TCC CTG TTG CTC Pro Ser Leu Leu Leu 5	TGG CCG CAT AGC CTT Trp Pro His Ser Leu 20	GGG CAG CGG TCT GTC Gly Gln Arg Ser Val 30	GTG CAT CCG GGT GCC Val His Pro Gly Ala 45	TTC ACC AAG TCA AAA Phe Thr Lys Ser Lys 60	AAC CCA GCC CTG CAT Asn Pro Ala Leu His 75
ATG Met 1	CTG Leu 15	ACA Thr	66C 61Y	TTT Phe	GGG G1Y
FIGURE 25 (I) CAAGGAGCCT CTGCCTGCCC GTCGTCGTC	G CCC CTG CGC CTA TGC CGG u Pro Leu Arg Leu Cys Arg 10	C CGA CTT CTC ACA GCC GCC e Arg Leu Leu Thr Ala Ala 25	T AAT TAC TAT GAA TTG TTG R ASN TYK TYK Glu Leu Leu 40	GCT GAA GAG ATT AAA CGT GCT Ala Glu Glu Ile Lys Arg Ala 50	A CAC CCT GAT CGA GAC CCT u His Pro Asp Arg Asp Pro 65
Figure Caagga	CAG CTG C	TCC ATC Ser Ile	CCT ACT Pro Thr 35	AGC GC Ser Al	GAG CTA Glu Leu

Substitute Sheet (Rule 26)

	6		383	425	467	209
	299	341	æ	4.2	4	Ŋ
		·				
	AGT Ser 90	TCA	AAG Lys	AAC Asn	GGG Gly	GTC Val 160
	CTC	CAT His	CCT Pro	CCC	CAG Gln	CGG
	GTG Val	CTG Leu	GAG Glu	CCC Pro 130	CCG Pro	CAG
	CGA Arg	CAG Gln	GCC Ala 115	GAA Glu	AGG Arg	AAC
	\mathtt{TAT}	CAC His 100	ACA Thr	TGG Trp	GTG Val	CAC His
	GCA Ala 85	GAC Asp	AGC Ser	TCC Ser	AGT Ser	AAA Lys 155
	GAG Glu	TAT	GGG G1y	AGC	CAC His	CGT Arg
	AAT	AAC Asn	TCA	AGC Ser 125	TTC Phe	CAG Gln
	CTG Leu	CGT Arg	TCT Ser 110	CAC His	CAG Gln	CAG Gln
	GAG Glu	CGT Arg 95	AAG Lys	ACA Thr	GCC Ala	AAG Lys
<u></u>	GTG Val 80	AGT Ser	CCA	CAG Gln	$^{ m TGG}$	AGG Arg 150
ר ר	TTT	GAA Glu	CCT Pro	CAA Gln	TAC	TCA
5 P G	AGC CGC TTT GT Ser Arg Phe Va	GAG Glu	AGT Ser	ACG Thr 120	CAA Gln	GAG Glu
ユート	AGC Ser	CGT	GCC Ala 105	TAT Tyr	GCT Ala	CCG Pro

81	185
UI	עט וו

CTC ATG	CTG CTC Leu Leu	CTC
AAG CTG	AGG AAG Arg Lys 180	AAG Lys
CGG ATC ATT Arg ile 1195	GAC CGG Asp Arg 195	CGG Arg 195
: AGG GCC AAC AGA GCC AGG ATT Arg Ala Asn Arg Ala Arg Ile 210	GCC AGG Ala Arg	AGG Arg
GAG CAG	AGG CAG Arg Gln	CAG Gln

FIGURE 25 (IV)	
CTG CCT CCA GAA AGC TCC AGG ATC ATG CCC CAG GAC ACA AGC Leu Pro Pro Glu Ser Ser Arg Ile Met Pro Gln Asp Thr Ser 240	761
CCC TGAGAGGCTT AACTAAATGG GACCTTCATT GGTCCTCTCC CTGCTGCTG Pro *	814
TCCAGAACTA CACGTGCAAT AAACTCATTT TCAG (A)n	849

human MCG18 MPPLL---PLRLCRLWPRNPPSRLLGAAAGQRSRPSTYYELLGVHPGASTEEVKRAFFSK MPSLLLQLPLRLCRLWPHSLSIRLLTAATGQRSVPTNYYELLGVHPGASAEEIKRAFFTK SKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRSYDDQLRSGSPPKSPRTTVHDKSA ******* human MCG18 mouse MCG18 mouse

FIGURE 26

SKELHPDRDPGNPALHSRFVELNEAYRVLSREESRRNYDHQLHSASPPKSSGSTAEPKYT ** ** ** *** ******* ****** ******* MCG18

human MCG18 HQTHSS-WTPPNAQYWSQFHSVRPQGPQLRQQQHKQNKQVLGYCLLLMLAGMGLHYIAFR QQTHSSSWEPPNAQYWAQFHSVRPQGPESRKQQRKHNQRVLGYCLLLMVAGMGLHYVAFR *** ****** ******* * ** * ******* ***** MCG18 mouse

KLEQVHRSFMDEKDRIITAIYNDTRARARANRARIQQER---HERQQPRAEPSLPPESSR KVKQMHLNFMDEKDRIITAFYNEARARARANRGILQQERQRLGQRQPPPSEPTQGPE--MCG18 mouse MCG18 human

** ********

human MCG18 IVPRGAGP

mouse MCG18 IMPQDTSP

Substitute Sheet (Rule 26)

40	8	120
GTA Val	SCC TTA Pro Leu 25	CC Pro
FC TTG eu Leu 10	CG CCC GCC ATG CCG CCC TI	AAC Asn
GCG CJ Ala Le	scc Arc	CCC CGC Pro Arg
c AAT r Asn	GCC GAIA	TGG Trp
TGG TC	GCC CO Pro 20	3G CTG rg Leu
ייייי	္လည္	TGC CGG Cys Arg
GCC CCA TCC Ala Pro Ser	CCC AGC TGC Pro Ser Cys	GC CTG TGC Arg Leu Cys
CTA GC Leu Al	T CCC	G CGC
AGT C Ser L	TCC TTT Ser Phe	CCC CTG
TTGA AGT * Ser	GCC 1 Ala S	CTG (Leu F

FIGURE 27

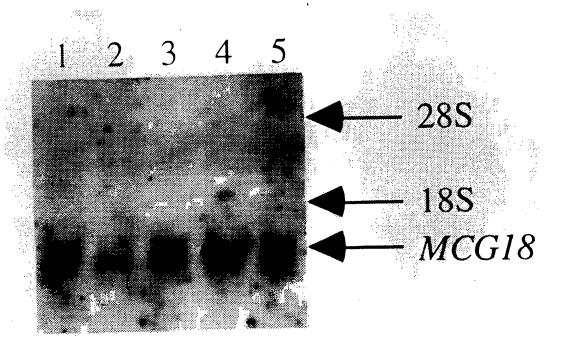


FIG 28

INTERNATIONAL SEARCH REPORT

International Application No. PCT/AU 98/00380

A.	CLASSIFICATION OF SUBJECT MATTER		
Int Cl ⁶ :	C12N 15/12; C07K 14/47; C07K 16/18; G01N 3	33/53	
According to	International Patent Classification (IPC) or to both	national classification and IPC	
В.	FIELDS SEARCHED		
Minimum doct	umentation searched (classification system followed by c WPAT (D gene) Sequences provided by Appli		
Documentation	n searched other than minimum documentation to the ext	tent that such documents are included in t	he fields searched
	a base consulted during the international search (name of nebank, Swiss Prot and PIR: Sequences provide :		terms used)
C.	DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where app	propriate, of the relevant passages	Relevant to claim No.
P,X	Kedra D, Seroussi E, Fransson I, Trifunovic Blennow E, Mehlin H, Dumanski J, Human G 611-619 The germinal centre kinase gene and located in the vicinity of the PYGM gene on EMBL AC Y12339	Genetics, October 1997 100(5-6) i a novel CDC25-like gene are	1-3,8-10,15-18
P,X	Guru S C, Agarwal S K, Manickain P, Olufe July 1997 7(7) 725-735. A transcript map for the multiple endocrine neoplasia type I locus TREMBL AC 014616		1. 4-5, 8, 11-12, 15, 19-21
X	Further documents are listed in the continuation of Box C	See patent family an	nex
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone document referring to an oral disclosure, use, exhibition or other means "P" document defining the general state of the art which is not considered to be of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone document referring to an oral disclosure, use, exhibition or other means "P" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art document member of the same patent family			
	ctual completion of the international search	Date of mailing of the international sear 2 0 JUL 1998	-
Name and ma AUSTRALIA PO BOX 200 WODEN AC AUSTRALIA	ailing address of the ISA/AU IN PATENT OFFICE CT 2606	Authorized officer	7(7116000

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 98/00380

Box 1	Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
	ernational Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following
1.	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	X Claims Nos.: 1, 2, 4, 6 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically: They are to known groups of proteins and lack distinguishing features which would enable a meaningful search.
3.	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)
Box II	Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
Inversions Inversions which Inversions which Inversions Inversion Inversi	International Searching Authority found multiple inventions in this international application, as follows: Intion 1, defined by claims 2, 3, 9, 10, 16-18, is to nucleotide sequences, amino acid sequences and proteins with a stringer domain. Intion 2, defined by claims 4, 5, 11, 12, 19-21, is to nucleotide sequences and amino acid sequences and proteins chare guanine exchange factors. Intion 3, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are heat shock proteins or heat shock binding proteins. Intion 3, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are heat shock proteins or heat shock binding proteins. Intion 3, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are heat shock proteins or heat shock binding proteins. Intion 2, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are intion 3, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are sequences and amino acid sequences and proteins are sequences and amino acid sequences and proteins are sequences. Intion 2, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are sequences and amino acid sequences and proteins are sequences. Intion 2, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are sequences. Intion 2, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are sequences. Intion 2, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are sequences. Intion 2, defined by claims 6, 7, 13, 14, 22-24, is to nucleotide sequences and amino acid sequences and proteins are sequences.
Rem	report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: ark on Protest The additional search fees were accompanied by the applicant's protest. X No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

international Application No.

PCT/AU 98/00380

ategory*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EMBL AC AF012106 DT 6 November 1997 Lloyd S E and Thakker R V DE Homo Sapiens DnaJ protein (HSPF ₂)mRNA, complete cds	1,6-8,13- 15,22-24
P,X	EMBL AC AF 036875 DT 20 May 1998 Silins G, Grimmond S, Hayward N DE Mus musculus multiple endocrine neoplasia type I candidate protein number 18 mRNA, complete cds	1,6-8,13- 15,22-24
•		
	Na.	

THIS PAGE BLANK (USP10)